



# **STIC Search Report**

## **Biotech-Chem Library**

**STIC Database Tracking Number: 99780**

**TO: Ginny Portner**  
**Location: CM1/7E13/7E12**  
**Art Unit: 1652**

July 30, 2003

**Case Serial Number: 09821348**

**From: P. Sheppard**  
**Location: CM1-1E03**  
**Phone: (703) 308-4499**

**sheppard@uspto.gov**

### **Search Notes**



## STIC-Biotech/ChemLib

---

**From:** Portner, Ginny  
**Sent:** Monday, July 28, 2003 9:39 AM  
**To:** STIC-Biotech/ChemLib  
**Subject:** 09/821,348

Please search SEQ ID NO 2 (peptide sequence) Thanks

*Ginny Portner*  
CM1, Art Unit 1645  
Room 7e13  
Mail box 7e12  
(703) 308-7543

Point of Contact  
P. Sheppard  
telephone number: (703) 308-4499  
7/30/03

Searcher: \_\_\_\_\_  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: \_\_\_\_\_  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

TYPE OF SEARCH:  
NA Sequences: \_\_\_\_\_  
AA Sequences: \_\_\_\_\_  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:45:27 ; Search time 95 Seconds  
(without alignments)  
62.476 Million cell updates/sec

Title: US-09-821-348-2  
Perfect score: 117  
Sequence: 1 VGVSIAGYQNGFTGNTTSAGP 23

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL 23:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_rvirus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	101	86.3	875	2	Q46221	Q46221 clostridium
2	79	67.5	876	2	O32739	O32739 clostridium
3	79	67.5	876	2	Q9KH41	Q9KH41 clostridium
4	72	61.5	879	2	O06498	O06498 clostridium
5	54	46.2	551	16	Q8XAQ0	Q8XAQ0 escherichia
6	54	46.2	551	16	Q8FBP5	Q8FBP5 escherichia
7	52	44.4	360	11	Q8K0A4	Q8K0A4 mus musculus
8	52	44.4	362	3	Q9HEZ9	Q9HEZ9 emericella
9	52	44.4	538	16	Q8U730	Q8U730 agrobacteri
10	52	44.4	852	11	Q8BZG1	Q8BZG1 mus musculus
11	51	43.6	689	11	Q8BSK0	Q8BSK0 mus musculus
12	51	43.6	804	11	Q99NC2	Q99NC2 mus musculus
13	51	43.6	906	11	Q99PM5	Q99PM5 mus musculus
14	51	43.6	965	11	Q99PW2	Q99PW2 mus musculus
15	51	43.6	1160	11	Q9WUN1	Q9WUN1 mus musculus
16	51	43.6	1160	11	Q8R564	Q8R564 mus musculus

17	51	43.6	1198	11	Q99PM6	Q99pm6 mus musculu
18	51	43.6	1198	11	Q924G8	Q924g8 mus musculu
19	51	43.6	1987	11	Q99PB3	Q99pb3 mus musculu
20	51	43.6	3600	10	Q9SA64	Q9sa64 arabidopsis
21	50.5	43.2	434	16	Q983K6	Q983k6 rhizobium l
22	50	42.7	129	16	Q8YQU6	Q8yqu6 anabaena sp
23	50	42.7	528	3	P87239	P87239 schizosacch
24	50	42.7	537	16	Q8FIZ4	Q8fiz4 escherichia
25	50	42.7	1341	16	Q8UAU1	Q8uau1 agrobacteri
26	49	41.9	357	16	Q9KU34	Q9ku34 vibrio chol
27	49	41.9	482	11	Q8BU52	Q8bu52 mus musculu
28	48.5	41.5	228	17	Q8PWM4	Q8pwm4 methanosarc
29	48	41.0	114	2	Q9RHG4	Q9rhg4 microcystis
30	48	41.0	160	4	Q8NC35	Q8nc35 homo sapien
31	48	41.0	228	11	Q8COL2	Q8col2 mus musculu
32	48	41.0	253	16	Q8ZQC5	Q8zqc5 salmonella
33	48	41.0	253	16	Q8Z804	Q8z804 salmonella
34	48	41.0	315	2	Q9X588	Q9x588 neisseria f
35	48	41.0	381	11	Q8BX42	Q8bx42 mus musculu
36	48	41.0	537	5	Q9Y154	Q9y154 drosophila
37	48	41.0	545	5	Q9VA35	Q9va35 drosophila
38	48	41.0	653	4	Q96JN4	Q96jn4 homo sapien
39	48	41.0	663	2	Q8RR60	Q8rr60 rhizobium s
40	48	41.0	727	4	Q9UDR4	Q9udr4 homo sapien
41	48	41.0	835	4	Q9ULF5	Q9ulf5 homo sapien
42	48	41.0	1323	2	O87018	O87018 helicobacte
43	48	41.0	1452	4	Q9H4A0	Q9h4a0 homo sapien
44	48	41.0	1512	4	Q9H4A1	Q9h4a1 homo sapien
45	47.5	40.6	3659	16	Q98LN6	Q98ln6 rhizobium l

ALIGNMENTS

RESULT 1

Q46221	ID	Q46221	PRELIMINARY;	PRT;	875 AA.
AC	Q46221;				
DT	01-NOV-1996	(TREMBLrel. 01, Created)			
DT	01-NOV-1996	(TREMBLrel. 01, Last sequence update)			
DT	01-JUN-2002	(TREMBLrel. 21, Last annotation update)			
DE	Iota toxin component Ib precursor.				
OS	Clostridium perfringens.				
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;				
OC	Clostridium.				
OX	NCBI_TaxID=1502;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=NCIB 10748;				
RX	MEDLINE=94041637; PubMed=8225592;				
RA	Perelle S., Gibert M., Boquet P., Popoff M.R.;				
RT	"Characterization of Clostridium perfringens iota toxin genes and expression in Escherichia coli.";				
RL	Infect. Immun. 61:5147-5156(1993).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=NCIB 10748;				
RA	Popoff M.R.;				
RL	Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; X73562; CAA51960.1; -.				
DR	HSSP; P13423; IACC.				
DR	InterPro; IPR003896; Anthrax toxinB.				
DR	pfam; PF03495; Binary toxB; 1.				
DR	PRINTS; PR01391; BINARYTOXINB.				
KW	Signal.				
FT	SIGNAL	34	38	POTENTIAL.	
FT	CHAIN	212	875	IOTA TOXIN COMPONENT IB.	
SQ	SEQUENCE	875 AA;	98468 MW;	C9AE092CD3818921 CRC64;	

Query Match 86.3%; Score 101; DB 2; Length 875;  
Best Local Similarity 100.0%; Pred. No. 1e-05;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VGVSISAGYQNGFTGNITTS 20  
||||:|||||:|||||  
Db 336 VGVSISAGYQNGFTGNITTS 355

## RESULT 2

O32739 PRELIMINARY; PRT; 876 AA.  
AC O32739;  
DT 01-JAN-1998 (TREMBlrel. 05, Created)  
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE ADP-ribosyltransferase.  
GN CDTB.  
OS Clostridium difficile.  
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
OC Clostridium.  
OX NCBI\_TaxID=1496;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CD196;  
RX MEDLINE=97230316; PubMed=9119480;  
RA Perelle S., Gilbert M., Bourlioux P., Corthier G., Popoff M.R.;  
RT "Production of a complete binary toxin (actin-specific ADP-  
RT ribosyltransferase) by Clostridium difficile CD196.";  
RL Infect. Immun. 65:1402-1407(1997).  
DR EMBL; L76081; AAB67305.1; -.  
DR HSSP; P13423; IACC.  
DR InterPro; IPR003896; Anthrax toxinB.  
DR Pfam; PF03495; Binary\_toxB; 1.  
DR PRINTS; PR01391; BINARYTOXINB.  
KW Transferase.  
SQ SEQUENCE 876 AA; 98797 MW; 25E06E2D45CE2B3B CRC64;

Query Match 67.5%; Score 79; DB 2; Length 876;  
Best Local Similarity 68.4%; Pred. No. 0.013;  
Matches 13; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 VGVSISAGYQNGFTGNITTS 20  
||||:|||||:|||||  
Db 338 GVSNNVGYQNGFTANVTN 356

## RESULT 3

Q9KH41 PRELIMINARY; PRT; 876 AA.  
AC Q9KH41;  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE CdtB.  
GN CDTB.  
OS Clostridium difficile.  
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
OC Clostridium.  
OX NCBI\_TaxID=1496;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CCUG 20309;  
RA Chang S.Y., Song K.P.;  
RT "ADP-ribosylating Binary Toxin Genes of Clostridium difficile strain  
RT CCUG 20309.";  
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF271719; AAF81761.1; -.  
DR HSSP; P13423; IACC.  
DR InterPro; IPR003896; Anthrax toxinB.  
DR Pfam; PF03495; Binary\_toxB; 1.  
DR PRINTS; PR01391; BINARYTOXINB.  
SQ SEQUENCE 876 AA; 98793 MW; 366D62F352E745A5 CRC64;

Query Match 67.5%; Score 79; DB 2; Length 876;  
Best Local Similarity 68.4%; Pred. No. 0.013;  
Matches 13; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 GVSISAGYQNGFTGNITTS 20  
||||:|||||:|||||  
Db 338 GVSNNVGYQNGFTANVTN 356

## RESULT 4

O06498 PRELIMINARY; PRT; 879 AA.  
AC O06498;  
DT 01-JUL-1997 (TREMBlrel. 04, Created)  
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Sb component.  
GN SBS.  
OS Clostridium spiroforme.  
OC Bacteria; Firmicutes; Mollicutes.  
OX NCBI\_TaxID=29348;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CS246;  
RA Gibert M., Perelle S., Daube G., Popoff M.R.;  
RT "Clostridium spiroforme toxin genes are related to C. perfringens iota  
RT toxin genes but have a different genomic localization.";  
RL Syst. Appl. Microbiol. 20:337-347(1997).  
DR EMBL; X97969; CAA66612.1; -.  
DR HSSP; P13423; IACC.  
DR InterPro; IPR003896; Anthrax toxinB.  
DR Pfam; PF03495; Binary\_toxB; 1.  
DR PRINTS; PR01391; BINARYTOXINB.  
SQ SEQUENCE 879 AA; 98739 MW; 40685ACB8E05BA01 CRC64;

Query Match 61.5%; Score 72; DB 2; Length 879;  
Best Local Similarity 68.4%; Pred. No. 0.12;  
Matches 13; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 GVSISAGYQNGFTGNITTS 20  
||||:|||||:|||||  
Db 341 GVAINIAYQNGFTGSITN 359

## RESULT 5

Q8XAQ0 PRELIMINARY; PRT; 551 AA.  
AC Q8XAQ0;  
DT 01-MAR-2002 (TREMBlrel. 20, Created)  
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
DE Arylsulfatase.  
GN ASLA OR Z5314 OR ECS4731.  
OS Escherichia coli O157:H7.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=83334;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;  
RX MEDLINE=21074935; PubMed=11206551;  
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouisis K.,  
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
RA Welch R.A., Blattner F.R.;  
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";  
RL Nature 409:529-533(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / RIMD 0509952;  
RX MEDLINE=21156231; PubMed=11258796;  
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,

```
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
DR EMBL; AE005611; AAG58993.1; -.
DR EMBL; AP002567; BAB38154.1; -.
DR InterPro; IPR000917; Sulfatase.
DR Pfam; PF00884; Sulfatase; 1.
DR PROSITE; PS00523; SULFATASE_1; 1.
DR PROSITE; PS00149; SULFATASE_2; 1.
KW Complete proteome.
SQ SEQUENCE 551 AA; 60654 MW; 449616F533D22DCF CRC64;

Query Match 46.2%; Score 54; DB 16; Length 551;
Best Local Similarity 56.2%; Pred. No. 25;
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 7 AGYQNGFTGNITTSAG 22
:||||| : :||
Db 486 SGYQGGFTGTVMQTAG 501

RESULT 6
Q8FBP5 PRELIMINARY; PRT; 551 AA.
AC Q8FBP5;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Arylsulfatase (EC 3.1.6.1).
GN ASLA OR C4719.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR EMBL; AE016769; AAN83152.1; -.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 551 AA; 60627 MW; A9819C4264B90C8A CRC64;

Query Match 46.2%; Score 54; DB 16; Length 551;
Best Local Similarity 56.2%; Pred. No. 25;
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 7 AGYQNGFTGNITTSAG 22
:||||| : :||
Db 486 SGYQGGFTGTVMQTAG 501

RESULT 7
Q8K0A4 PRELIMINARY; PRT; 360 AA.
AC Q8K0A4;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Similar to cell division cycle 2-like 5, isoform 1, cholinesterase-
DE related cell division controller, CDC2-related protein kinase 5
DE (Fragment).
GN CDC2L5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```

```
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast tumor;
RA Strausberg R.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC032179; AAH32179.1; -.
DR MGD; MGI:1916812; Cdc215.
KW Cell division.
FT NON TER 1
SQ SEQUENCE 360 AA; 38609 MW; A20848BDEA15716C CRC64;

Query Match 44.4%; Score 52; DB 11; Length 360;
Best Local Similarity 60.0%; Pred. No. 30;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 8 GYQNGFTGNITTSAG 22
||| :|||
Db 337 GYSQGYRGHISTSAG 351

RESULT 8
Q9HEZ9 PRELIMINARY; PRT; 362 AA.
AC Q9HEZ9;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Tyr-inhibited DAHP synthase (EC 4.1.2.15).
GN AROF.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=162425;
RN [1]
RP SEQUENCE FROM N.A.
RA Hartmann M., Heinrich G., Braus G.H.;
RT "Two novel genes arof and aroG of Aspergillus nidulans code for
RT differently regulated DAHP synthases.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF283008; AAG36949.1; -.
DR HSSP; P00886; 1QR7.
DR InterPro; IPR006219; AROFGH.
DR InterPro; IPR006218; DAHP1/KDSA.
DR Pfam; PF00793; DAHP synth 1; 1.
DR ProDom; PD005060; AROFGH_1.
DR TIGRFAMs; TIGR00034; aroFGH; 1.
KW Lyase.
SQ SEQUENCE 362 AA; 39260 MW; 6DB9AADF35D3280B CRC64;

Query Match 44.4%; Score 52; DB 3; Length 362;
Best Local Similarity 47.4%; Pred. No. 30;
Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 2 GVSISAGYQNGFTGNITTS 20
||| :|||
Db 178 GLSPPIGYKNGTDGNLTVA 196

RESULT 9
Q8U730 PRELIMINARY; PRT; 538 AA.
AC Q8U730;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE ABC transporter, substrate binding protein.
GN ATU4626 OR AGR L 514.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=176299;
RN [1]
```

RP SEQUENCE FROM N.A.  
RX MEDLINE=21608550; PubMed=11743193;  
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,  
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,  
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,  
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,  
RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,  
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,  
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Pery M.,  
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Doian M.,  
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,  
RA Neeter E.W.;  
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens  
RT C58.";  
RL Science 294:2317-2323(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21608551; PubMed=11743194;  
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,  
RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,  
RA Houniel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,  
RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,  
RA Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,  
RA Cielo C., Slater S.;  
RT "Genome sequence of the plant pathogen and biotechnology agent  
RT Agrobacterium tumefaciens C58.";  
RL Science 294:2323-2328(2001).  
DR EMBL; AE009390; AAL45420.1; ALT\_INIT.  
DR EMBL; AE008225; AAK88821.1; -.  
BR InterPro; IPR000914; SBP\_bac\_5.  
DR Pfam; PF00496; SBP\_bac\_5; 1.  
KW Complete proteome.  
SQ SEQUENCE 538 AA; 59011 MW; 1223C55B1DC1537C CRC64;  
  
Query Match 44.4%; Score 52; DB 16; Length 538;  
Best Local Similarity 66.7%; Pred. No. 47;  
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
  
QY 7 AGYQNGFTGNITTSAG 21  
Db |||:|||||:|  
366 AGYENGFTLNLTIA 380  
  
RESULT 10  
Q8BZG1 ID Q8BZG1 PRELIMINARY; PRT; 852 AA.  
AC Q8BZG1;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Cell division cycle 2-like 5.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=C57BL/6J; TISSUE=Urinary bladder;  
RX MEDLINE=22354683; PubMed=12466851;  
RA The FANTOM Consortium,  
RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs.";  
RL Nature 420:563-573(2002).  
DR EMBL; AK035493; BAC29077.1; -.  
SQ SEQUENCE 852 AA; 94693 MW; COD17E973798A379 CRC64;  
  
Query Match 44.4%; Score 52; DB 11; Length 852;  
Best Local Similarity 60.0%; Pred. No. 77;  
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
  
QY 8 GYQNGFTGNITTSAG 22  
Db |||:|||||

Db 829 GYSQGYRGHISTSAG 843  
  
RESULT 11  
Q8BSK0 ID Q8BSK0 PRELIMINARY; PRT; 689 AA.  
AC Q8BSK0;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Melanoma antigen (Fragment).  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=C57BL/6J; TISSUE=Mesonephros;  
RX MEDLINE=22354683; PubMed=12466851;  
RA The FANTOM Consortium,  
RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs.";  
RL Nature 420:563-573(2002).  
DR EMBL; AK032791; BAC28024.1; -.  
FT NON TER 1  
SQ SEQUENCE 689 AA; 62486 MW; 9E8D8D7EDEE1A78F CRC64;  
  
Query Match 43.6%; Score 51; DB 11; Length 689;  
Best Local Similarity 42.9%; Pred. No. 84;  
Matches 12; Conservative 3; Mismatches 7; Indels 6; Gaps 1;  
  
QY 2 GVSISAGY-----QNGFTGNITTSAGF 23  
Db |||:|||||  
636 GPSTAAGFGSGLSTSTGTGFGGLNTSAGF 663  
  
RESULT 12  
Q99NC2 ID Q99NC2 PRELIMINARY; PRT; 804 AA.  
AC Q99NC2;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE MAGE-necdin/trophinin complex (Magphinin).  
GN TRO OR MAGED3 OR TROPHININ/MAGPHININ.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Saburi S., Hirama K., Yamanouchi K., Naito K., Tojo H., Tachi C.,  
RA Fukuda M.;  
RT "Identification of an Isoform of Trophinin Gene Transcript Coding for  
RT a Novel Protein, Magphinin, Revealed the Complex Genomic Structure of  
RT the Trophinin Gene.";  
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB032477; BAB40318.1; -.  
DR MGD; MGI:1928994; Tro.  
DR InterPro; IPR002190; MAGE.  
DR Pfam; PF01454; MAGE; 1.  
DR PROSITE; PS50838; MAGE; 1.  
SQ SEQUENCE 804 AA; 85038 MW; 3548B61E33F915AE CRC64;  
  
Query Match 43.6%; Score 51; DB 11; Length 804;  
Best Local Similarity 42.9%; Pred. No. 1e+02;  
Matches 12; Conservative 3; Mismatches 7; Indels 6; Gaps 1;  
  
QY 2 GVSISAGY-----QNGFTGNITTSAGF 23  
Db |||:|||||  
751 GPSTAAGFGSGLSTSTGTGFGGLNTSAGF 778







GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:38:32 ; Search time 23 Seconds  
(without alignments)  
47.027 Million cell updates/sec

Title: US-09-821-348-2  
Perfect score: 117  
Sequence: 1 VGVISAGYQNGFTGNITTSAGP 23

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues  
Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	54	46.2	551	1 ASLA_ECOLI	P25549 escherichia
2	53	45.3	1387	1 TROP_HUMAN	Q12816 homo sapien
3	46	39.3	212	1 MSRA_VIBCH	Q9kp30 vibrio chol
4	45	38.5	73	1 IAA1_STROI	P09921 streptomyc
5	45	38.5	75	1 IAA2_STROI	P20596 streptomyc
6	45	38.5	356	1 AROF_ECOLI	P00888 escherichia
7	45	38.5	356	1 AROF_SALTY	P21307 salmonella
8	45	38.5	359	1 Y130_ARCFU	Q30107 archaeglob
9	45	38.5	366	1 AROG_CORGL	P35170 corynebacte
10	45	38.5	483	1 CORT_DROME	Q960n3 drosophila
11	45	38.5	1276	1 PMP6_CHLPN	Q9z899 chlamydia p
12	44.5	38.0	212	1 Y440_CHLPN	Q9z8a3 chlamydia p
13	44.5	38.0	250	1 TIP1_TOBAC	P21653 nicotiana t
14	44.5	38.0	250	1 TIP2_TOBAC	P24422 nicotiana t
15	44.5	38.0	399	1 K1CS_BOVIN	P08728 bos taurus
16	44.5	38.0	943	1 ODO1_AZOVI	P20707 azotobacter
17	44	37.6	238	1 OP68_NEIGO	Q04881 neisseria g
18	44	37.6	400	1 K1CS_HUMAN	P08727 homo sapien
19	44	37.6	431	1 ENO_SHEON	Q8ebro shewanella
20	44	37.6	457	1 PRTC_STRGR	P52320 streptomyc
21	44	37.6	637	1 GYRB_BACHD	O50627 bacillus ha
22	44	37.6	868	1 MCM2_YEAST	P29469 saccharomyc
23	43	36.8	213	1 AROF_ERWHE	Q02285 erwinia her
24	43	36.8	233	1 OP67_NEIGO	Q05034 neisseria g
25	43	36.8	238	1 OP66_NEIGO	Q05033 neisseria g
26	43	36.8	368	1 ALR3_SALTI	Q8z300 salmonella
27	43	36.8	546	1 FLGK_ECOLI	P32335 escherichia
28	43	36.8	649	1 PTWA_BUCAP	Q8k911 buchnera ap
29	43	36.8	793	1 D153_HAEIN	O32629 haemophilus
30	43	36.8	795	1 D152_HAEIN	P44935 haemophilus
31	43	36.8	797	1 D151_HAEIN	P46024 haemophilus
32	43	36.8	1310	1 VAC3_HELPY	Q48253 helicobacte
33	42	35.9	104	1 Y4EB_RHISN	P55425 rhizobium s

34	42	35.9	104	1 YJA7_YEAST	P47080 saccharomyc
35	42	35.9	240	1 LEC_LOTTE	P19664 lotus tetra
36	42	35.9	447	1 VID2_AGR5	P18592 agrobacteri
37	42	35.9	534	1 HUP1_CHLKE	P15686 chlorella k
38	42	35.9	556	1 FTHS_CLOCY	Q07064 clostridium
39	42	35.9	750	1 CBBB_BACTV	Q9zius bacillus th
40	42	35.9	826	1 YEHB_ECOLI	P33341 escherichia
41	42	35.9	886	1 MCM2_XENLA	P55861 xenopus lae
42	42	35.9	895	1 MCM2_HUMAN	P49736 homo sapien
43	42	35.9	904	1 MCM2_MOUSE	P97310 mus musculu
44	42	35.9	1006	1 BGAL_ASPNG	P29853 aspergillus
45	41.5	35.5	396	1 PGK_STAAM	Q925c4 staphylococ

ALIGNMENTS

RESULT 1  
ASLA\_ECOLI  
ID ASLA\_ECOLI STANDARD; PRT; 551 AA.  
AC P25549;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Arylsulfatase (EC 3.1.1.6.1) (Aryl-sulfate sulphonylase).  
GN ASLA OR ATSA OR B3801.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=92358234; PubMed=1379743;  
RA Daniels D.L., Plunkett G. III, Burland V.D., Blattner F.R.;  
RT "Analysis of the Escherichia coli genome: DNA sequence of the region  
from 84.5 to 86.5 minutes."  
RL Science 257:771-778(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RA Murphy H.R., Kalman M., Cashel M.;  
RL Submitted (APR-1992) to the EMBL/GenBank/DBJ databases.  
CC -! CATALYTIC ACTIVITY: A phenol sulfate + H(2)O = a phenol + sulfate.  
CC -! SIMILARITY: BELONGS TO THE SULFATASE FAMILY.  
CC -! CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 462  
ONWARD AND IS SHORTER (475 AA) DUE TO A FRAMESHIFT.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).  
-----  
CC EMBL; M87049; AAA67597.1; -.  
DR EMBL; AE000456; AAC76804.1; -.  
DR EMBL; M90498; AAC32036.1; ALT\_FRAME.  
DR PIR; S30691; S30691.  
DR HSSP; P15848; 1FSU.  
DR EcoGene; EG10089; asla.  
DR InterPro; IPR000917; Sulfatase.  
DR Pfam; PF00884; Sulfatase; 1.  
DR PROSITE; PS00523; SULFATASE\_1; 1.  
DR PROSITE; PS00149; SULFATASE\_2; 1.  
KW Hydrolase; Complete proteome.  
FT ACT\_SITE 190 190 POTENTIAL.  
FT CONFLICT 191 191 M -> I (IN REF. 2).  
SQ SEQUENCE 551 AA; 60717 MW; A04C8BBE5E63E988 CRC64;

Query Match 46.2%; Score 54; DB 1; Length 551;  
Best Local Similarity 56.2%; Pred. No. 1.7;

Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 7 AGYQNGFTGNITTSAG 22  
:|||||: :||  
Db 486 SGYQGGFTGTVMQTAG 501

RESULT 2  
TROP HUMAN STANDARD; PRT; 1387 AA.  
AC Q12816; Q9NU89; Q9UPN8;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Trophinin.  
GN TRO OR KIAA1114.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=99397452; PubMed=10470851;  
RA Kikuno R., Nagase T., Ishikawa K.-I., Hirose M., Miyajima N.,  
Tanaka A., Kotani H., Nomura N., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes. XIV.  
RT The complete sequences of 100 new cDNA clones from brain which code  
RT for large proteins in vitro.";  
RL DNA Res. 6:197-205(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Ambrose K.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 639-1387 FROM N.A.  
RX MEDLINE=95278733; PubMed=7758945;  
RA Fukuda M.N., Sato T., Nakayama J., Klier G., Mikami M., Aoki D.,  
Nozawa S.;  
RT "Trophinin and tastin, a novel cell adhesion molecule complex with  
RT potential involvement in embryo implantation.";  
RL Genes Dev. 9:1199-1210(1995).  
CC -|- FUNCTION: COULD BE INVOLVED WITH BYSTIN AND TASTIN IN A CELL  
CC ADHESION MOLECULE COMPLEX THAT MEDIATES AN INITIAL ATTACHMENT OF  
CC THE BLASTOCYST TO UTERINE EPITHELIAL CELLS AT THE TIME OF THE  
CC EMBRYO IMPLANTATION. DIRECTLY RESPONSIBLE FOR HOMOPHILIC CELL  
CC ADHESION.  
CC -|- SUBUNIT: DIRECTLY BINDS BYSTIN, AND INDIRECTLY TASTIN.  
CC -|- TISSUE SPECIFICITY: STRONG EXPRESSION AT IMPLANTATION SITES. FOUND  
CC IN THE PLACENTA FROM THE SIXTH WEEK OF PREGNANCY. WAS LOCALIZED IN  
CC THE CYTOPLASM OF THE SYNCYTOTROPHOBLAST IN THE CHORIONIC VILLI  
CC AND IN ENDOMETRIAL DECIDUAL CELLS AT THE UTEROPLACENTAL INTERFACE.  
CC AFTER WEEK 10, THE LEVEL DECREASED AND THEN DISAPPEARED FROM  
CC PLACENTAL VILLI. ALSO FOUND IN MACROPHAGES.  
CC -|- SIMILARITY: Contains 1 MAGE domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AB029037; BAA83066.1; -.  
DR EMBL; AL049732; CAB86651.1; -.  
DR EMBL; U04811; AAA79334.1; -.  
DR PIR; I38488; I38488.  
DR Genew; HGNC:12326; TRO.  
DR MIM; 300132; -.  
DR GO; GO:0005887; C:integral to plasma membrane; TAS.  
DR GO; GO:0007566; P:embryo implantation; TAS.  
DR GO; GO:0007156; P:homophilic cell adhesion; TAS.

DR InterPro; IPR002190; MAGE.  
DR Pfam; PF01454; MAGE; 1.  
DR PROSITE; PS50838; MAGE; 1.  
KW Cell adhesion; Antigen; Repeat.  
FT DOMAIN 400 598 MAGE.  
FT DOMAIN 707 1381 69 X 10 AA APPROXIMATE TANDEM REPEATS.  
FT CONFLICT 653 653 A -> S (IN REF. 3).  
FT CONFLICT 692 692 G -> R (IN REF. 3).  
FT CONFLICT 694 694 S -> G (IN REF. 1).  
FT CONFLICT 724 724 S -> G (IN REF. 1).  
FT CONFLICT 768 768 S -> G (IN REF. 1).  
FT CONFLICT 780 780 S -> G (IN REF. 1).  
FT CONFLICT 982 982 N -> G (IN REF. 1).  
FT CONFLICT 1012 1012 S -> G (IN REF. 1).  
SQ SEQUENCE 1387 AA; 138804 MW; 412DIE5754B5A416 CRC64;

Query Match 45.3%; Score 53; DB 1; Length 1387;  
Best Local Similarity 39.3%; Pred. No. 6.2;  
Matches 11; Conservative 7; Mismatches 4; Indels 6; Gaps 1;

QY 2 GVSISAGY-----QNGFTGNITTSAGF 23  
|::|||: |||:|:|:|:  
Db 1168 GLNTSAGFGGGLGTSGAGFSGGLSTSSGF 1195

RESULT 3  
MSRA VIBCH STANDARD; PRT; 212 AA.  
AC Q9KP30;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Peptide methionine sulfoxide reductase msra (EC 1.8.4.6) (Protein-  
DE methionine-S-oxide reductase) (Peptide Met(O) reductase).  
GN MSRA OR VC2549.  
OS Vibrio cholerae.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
OC Vibrionaceae; Vibrio.  
OX NCBI\_TaxID=666;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=El Tor N16961 / Serotype O1;  
RX MEDLINE=20406833; PubMed=10952301;  
RA Heidelberg J.F., Eissen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,  
Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,  
Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,  
McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,  
Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
Fraser C.M.;  
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio  
RT cholerae.";  
RL Nature 406:477-483(2000).  
CC -|- FUNCTION: Has an important function as a repair enzyme for  
CC proteins that have been inactivated by oxidation. Catalyzes the  
CC reversible oxidation-reduction of methionine sulfoxide in proteins  
CC to methionine (By similarity).  
CC -|- CATALYTIC ACTIVITY: Protein L-methionine + oxidized thioredoxin =  
CC protein L-methionine S-oxide + reduced thioredoxin.  
CC -|- SIMILARITY: BELONGS TO THE MSRA MET SULFOXIDE REDUCTASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AE004324; AAF95690.1; -.  
DR PIR; E82061; E82061.  
DR HSSP; P54149; 1FVA.  
DR TIGR; VC2549; -.

```

DR HAMAP; MF 01401; -; 1.
DR InterPro; IPR002569; PMSR.
DR Pfam; PF01625; PMSR; 1.
DR ProDom; PD003489; PMSR; 1.
DR TIGRFAMs; TIGR00401; msrA; 1.
KW Oxidoreductase; Complete proteome.
FT ACT SITE 51 51 BY SIMILARITY.
SQ SEQUENCE 212 AA; 23360 MW; AD50686E8BF13016 CRC64;

Query Match 39.3%; Score 46; DB 1; Length 212;
Best Local Similarity 56.2%; Pred. No. 9.6;
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 3 VSISAGYQNGFTGNIT 18
DB 66 ISTSVGVSGGFTPNPT 81

RESULT 4
ID IAA1_STROI STANDARD; PRT; 73 AA.
AC P09921;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-FEB-1991 (Rel. 17, Last annotation update)
DE Alpha-amylase inhibitor PAIM I (PIG pancreatic alpha-amylase inhibitor
DE of microbes I).
OS Streptomyces olivaceoviridis (Streptomyces corchorusii).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1921;
RN [1]
RP SEQUENCE.
RX MEDLINE=88107556; PubMed=3501315;
RA Hirayama K., Takahashi R., Akashi S., Fukuhara K., Oouchi N.,
RA Murai A., Arai M., Murao S., Tanaka K., Nojima I.;
RT "Primary structure of Paim I, an alpha-amylase inhibitor from
RT Streptomyces corchorushii, determined by the combination of Edman
RT degradation and fast atom bombardment mass spectrometry.";
RL Biochemistry 26:6483-6488(1987).
CC -|- FUNCTION: INHIBITS MAMMALIAN ALPHA-AMYLASES SPECIFICALLY BUT
CC HAS NO ACTION ON PLANT AND MICROBIAL ALPHA-AMYLASES.
CC -|- MISCELLANEOUS: PAIM I INHIBITS ALPHA-AMYLASES FROM PIG, DOG, COW,
CC HORSE, BUT HAS NO ACTIVITY AGAINST HUMAN SALIVARY AND PANCREATIC
CC AMYLASES.
DR HSSP; P01092; 1BVN.
DR InterPro; IPR000833; A_amylase_inhib.
DR Pfam; PF01356; A_amylase_inhib; 1.
DR ProDom; PD009058; A_amylase_inhib; 1.
KW Alpha-amylase inhibitor.
FT DISULFID 8 24 BY SIMILARITY.
FT DISULFID 42 70 BY SIMILARITY.
SQ SEQUENCE 73 AA; 7424 MW; 0AB2E59906192C3B CRC64;

Query Match 38.5%; Score 45; DB 1; Length 73;
Best Local Similarity 52.9%; Pred. No. 4.5;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 3 VSISAGYQNGFTGNITT 19
DB 28 VSVSVAYQDGATGPCAT 44

RESULT 5
ID IAA2_STROI STANDARD; PRT; 75 AA.
AC P20596;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-FEB-1991 (Rel. 17, Last annotation update)
DE Alpha-amylase inhibitor PAIM II (PIG pancreatic alpha-amylase
DE inhibitor of microbes II).
OS Streptomyces olivaceoviridis (Streptomyces corchorusii).

```

```

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1921;
RN [1]
RP SEQUENCE.
RX MEDLINE=89134256; PubMed=2783847;
RA Akashi S., Hirayama K., Murai A., Arai M., Murao S.;
RT "Determination of the primary structure of Paim II, an alpha-amylase
RT inhibitor from Streptomyces corchorushii, by high-performance tandem
RT mass spectrometry.";
RL Biochem. Biophys. Res. Commun. 158:514-519(1989).
CC -|- FUNCTION: INHIBITS MAMMALIAN ALPHA-AMYLASES SPECIFICALLY BUT
CC HAS NO ACTION ON PLANT AND MICROBIAL ALPHA-AMYLASES.
DR PIR; A31298; A31298.
DR HSSP; P01092; 1BVN.
DR InterPro; IPR000833; A_amylase_inhib.
DR Pfam; PF01356; A_amylase_inhib; 1.
DR ProDom; PD009058; A_amylase_inhib; 1.
KW Alpha-amylase inhibitor.
FT DISULFID 10 26 BY SIMILARITY.
FT DISULFID 44 72 BY SIMILARITY.
SQ SEQUENCE 75 AA; 7626 MW; 0AE6A3548D5FE1E4 CRC64;

Query Match 38.5%; Score 45; DB 1; Length 75;
Best Local Similarity 52.9%; Pred. No. 4.6;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 3 VSISAGYQNGFTGNITT 19
DB 30 VSVSVAYQDGATGPCAT 46

RESULT 6
AROF_ECOLI STANDARD; PRT; 356 AA.
ID AROF_ECOLI STANDARD; PRT; 356 AA.
AC P00888;
DT 21-JUL-1986 (Rel. 01, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phospho-2-dehydro-3-deoxyheptonate aldolase, Tyr-sensitive
DE (EC 4.1.2.15) (Phospho-2-keto-3-deoxyheptonate aldolase) (DAHP
DE synthetase) (3-deoxy-D-arabino-heptulosonate 7-phosphate synthase).
GN AROF OR B2601.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84264621; PubMed=6146618;
RA Shultz J., Hermodson M.A., Garner C.C., Herrmann K.M.;
RT "The nucleotide sequence of the aroF gene of Escherichia coli and the
RT amino acid sequence of the encoded protein, the tyrosine-sensitive 3-
RT deoxy-D-arabino-heptulosonate 7-phosphate synthase.";
RL J. Biol. Chem. 259:9655-9661(1984).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=85134883; PubMed=6396419;
RA Hudson G.S.;
RT "Nucleotide sequence and transcription of the phenylalanine and
RT tyrosine operons of Escherichia coli K12.";
RL J. Mol. Biol. 180:1023-1051(1984).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [4]

```







```
CC -----
DR EMBL; L07603; AAA23292.1; -.
DR EMBL; AP005277; BAB98383.1; -.
DR PIR; I40637; I40837.
DR HSSP; P00886; IQR7.
DR InterPro; IPR006219; AroFGH.
DR InterPro; IPR006218; DAHP1/KDSA.
DR Pfam; PF00793; DAHP_synth_1; 1.
DR ProDom; PD005060; AroFGH; 1.
DR TIGRFAMs; TIGR00034; aroFGH; 1.
KW Aromatic amino acid biosynthesis; Lyase; Complete proteome.
FT CONFLICT 109 109 I -> T (IN REF. 1).
FT CONFLICT 181 181 R -> RR (IN REF. 1).
FT CONFLICT 356 356 A -> T (IN REF. 1).
SQ SEQUENCE 366 AA; 39131 MW; 1EDA38DD26E29D58 CRC64;

Query Match 38.5%; Score 45; DB 1; Length 366;
Best Local Similarity 50.0%; Pred. No. 24;
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 GVSISAGYQNGFTGNI 17
Db 186 GMSMPIGPKNGTGDGNI 201

RESULT 10
CORT DROME STANDARD; PRT; 483 AA.
ID CORT DROME STANDARD; PRT; 483 AA.
AC Q960N3; Q9VMA1;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Protein cortex.
GN CORT OR CG11330.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, TISSUE
RP SPECIFICITY, AND MUTANTS QW55 AND RH65.
RX MEDLINE=21150807; PubMed=11252055;
RA Chu T., Henrion G., Haegeli V., Strickland S.;
RT "Cortex, a Drosophila gene required to complete oocyte meiosis, is a
RT member of the Cdc20/fizzy protein family.";
RL Genesis 29:141-152(2001).
RN [2]
RP SEQUENCE FROM N.A.
RP STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foslter C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
```

```
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [3]
RP SEQUENCE OF 10-483 FROM N.A.
RC STRAIN=Berkeley; TISSUE=Embryo;
RX MEDLINE=22426066; PubMed=12537569;
RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA George R.A., Guarin H., Kronmiller B., Pacleb J.M., Park S., Wan K.H.,
RA Rubin G.M., Celniker S.E.;
RT "A Drosophila full-length cDNA resource.";
RL Genome Biol. 3:RESEARCH080.1-RESEARCH080.8(2002).
RN [4]
RP INDUCTION.
RX MEDLINE=20384592; PubMed=10924478;
RA Harms E., Chu T., Henrion G., Strickland S.;
RT "The only function of Grauzone required for Drosophila oocyte meiosis
RT is transcriptional activation of the cortex gene.";
RL Genetics 155:1831-1839(2000).
CC -!- FUNCTION: Essential for female meiosis.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- TISSUE SPECIFICITY: Highly expressed in ovaries. Expressed in
CC nurse cells from stage 6 to stage 12 egg chambers. As oocyte
CC matures, it is transferred from the nurse cells to the oocyte.
CC -!- DEVELOPMENTAL STAGE: Zygotically expressed during oogenesis and
CC maternally deposited in oocytes. Maternal expression rapidly drops
CC off after 1hr.
CC -!- INDUCTION: Expression is transcriptionally activated by Grau.
CC -!- SIMILARITY: Contains 5 WD repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AY033478; AAK54464.1; -.
CC EMBL; AE003614; AAF52421.1; -.
CC EMBL; AY051966; AAK93390.1; ALT_INIT.
CC FlyBase; FBgn0000351; cort.
CC InterPro; IPR001680; WD40.
CC Pfam; PF00400; WD40; 3.
CC PROSITE; PS00678; WD_REPEATS_1; FALSE_NEG.
CC PROSITE; PS50082; WD_REPEATS_2; 2.
CC PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW Meiosis; Repeat; Developmental protein; WD repeat.
FT REPEAT 158 195 WD 1.
FT REPEAT 196 237 WD 2.
FT REPEAT 245 284 WD 3.
FT REPEAT 287 326 WD 4.
FT REPEAT 421 460 WD 5.
FT MUTAGEN 303 303 Y->C: IN QW55; ABNORMAL ARREST IN FEMALE
FT MEIOSIS.
FT MUTAGEN 431 483 MISSING: IN RH65; ABNORMAL ARREST IN
FT FEMALE MEIOSIS.
SQ SEQUENCE 483 AA; 55574 MW; 007DD0190DE06735 CRC64;

Query Match 38.5%; Score 45; DB 1; Length 483;
```





DR PIR; F72077; F72077.  
DR PIR; F86545; F86545.  
DR TIGR; CP0313; -.  
KW Hypothetical protein; Transmembrane; Complete proteome.  
FT TRANSMEM 54 74 POTENTIAL.  
FT TRANSMEM 79 99 POTENTIAL.  
SQ SEQUENCE 212 AA; 23251 MW; 719F48EB322875B1 CRC64;

Query Match 38.0%; Score 44.5; DB 1; Length 212;  
Best Local Similarity 62.5%; Pred. No. 16;  
Matches 10; Conservative 3; Mismatches 2; Indels 1; Gaps 1;

QY 1 VGVSIISAGYQNGFTGN 16  
:|:||||| |:|  
Db 64 LGGTISAGYA-GYTGN 78

RESULT 13  
TIP1\_TOBAC  
ID TIP1\_TOBAC STANDARD; PRT; 250 AA.  
AC P21653;  
DT 01-MAY-1991 (Rel. 18, Created)  
DT 01-MAY-1991 (Rel. 18, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Tonoplast intrinsic protein, root-specific RB7-5A (RT-TIP).  
OS Nicotiana tabacum (Common tobacco).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; lamiids; Solanales; Solanaceae; Nicotiana.  
OX NCBI\_TaxID=4097;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93005705; PubMed=1840917;  
RA Yamamoto Y.T., Taylor C.G., Acedo G.N., Cheng C.-L., Conkling M.A.;  
RT "Characterization of cis-acting sequences regulating root-specific  
gene expression in tobacco.";  
RL Plant Cell 3:371-382(1991).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Wisconsin 38; TISSUE=Root;  
RX MEDLINE=91081344; PubMed=2129561;  
RA Yamamoto Y.T., Cheng C.-L., Conkling M.A.;  
RT "Root-specific genes from tobacco and Arabidopsis homologous to an  
evolutionarily conserved gene family of membrane channel proteins.";  
RL Nucleic Acids Res. 18:7449-7449(1990).  
CC -!- FUNCTION: CHANNEL PROTEIN IN TONOPLAST. THESE PROTEINS MAY ALLOW  
THE DIFFUSION OF AMINO ACIDS AND/OR PEPTIDES FROM THE TONOPLAST  
INTERIOR TO THE CYTOPLASM.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -!- TISSUE SPECIFICITY: Roots.  
CC -!- SIMILARITY: BELONGS TO THE MIP/AQUAPORIN FAMILY (TC 1.A.8).

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

EMBL; S45406; AAB23597.2; -.  
EMBL; X54855; CAA38634.1; -.  
DR PIR; S13719; S13719.  
DR HSSP; P29972; 1H61.  
DR InterPro; IPR000425; MIP\_family.  
DR Pfam; PF00230; MIP; 1.  
DR PRINTS; PR00783; MINTRINSICP.  
DR ProDom; PD000295; MIP family; 1.  
DR TIGRFAMs; TIGR00861; MIP; 1.  
DR PROSITE; PS00221; MIP; 1.  
KW Transport; Transmembrane; Multigene family.  
SQ SEQUENCE 250 AA; 25233 MW; 6B243E8EB1DB0C1D CRC64;

Query Match 38.0%; Score 44.5; DB 1; Length 250;  
Best Local Similarity 38.2%; Pred. No. 19;  
Matches 13; Conservative 1; Mismatches 9; Indels 11; Gaps 1;  
QY 1 VGVSIISAGYQNGFT-----GNITTSAGF 23  
|||||:| |  
Db 69 VGVSIAANISGGHLPVTLGLAVGGNITLTGF 102

RESULT 14  
TIP2\_TOBAC  
ID TIP2\_TOBAC STANDARD; PRT; 250 AA.  
AC P24422;  
DT 01-MAR-1992 (Rel. 21, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 01-JUN-1994 (Rel. 29, Last annotation update)  
DE Tonoplast intrinsic protein, root-specific RB7-18C (RT-TIP).  
OS Nicotiana tabacum (Common tobacco).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; lamiids; Solanales; Solanaceae; Nicotiana.  
OX NCBI\_TaxID=4097;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Root;  
RX MEDLINE=93005705; PubMed=1840917;  
RA Yamamoto Y.T., Taylor C.G., Acedo G.N., Cheng C.-L., Conkling M.A.;  
RT "Characterization of cis-acting sequences regulating root-specific  
gene expression in tobacco.";  
RL Plant Cell 3:371-382(1991).  
CC -!- FUNCTION: CHANNEL PROTEIN IN TONOPLAST. THESE PROTEINS MAY ALLOW  
THE DIFFUSION OF AMINO ACIDS AND/OR PEPTIDES FROM THE TONOPLAST  
INTERIOR TO THE CYTOPLASM.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -!- TISSUE SPECIFICITY: Roots.  
CC -!- SIMILARITY: BELONGS TO THE MIP/AQUAPORIN FAMILY (TC 1.A.8).  
DR PIR; JQ1012; JQ1012.  
DR HSSP; P29972; 1H61.  
DR InterPro; IPR000425; MIP\_family.  
DR Pfam; PF00230; MIP; 1.  
DR PRINTS; PR00783; MINTRINSICP.  
DR ProDom; PD000295; MIP family; 1.  
DR TIGRFAMs; TIGR00861; MIP; 1.  
DR PROSITE; PS00221; MIP; 1.  
KW Transport; Transmembrane; Multigene family.  
SQ SEQUENCE 250 AA; 25224 MW; 2BFF533547006438 CRC64;

Query Match 38.0%; Score 44.5; DB 1; Length 250;  
Best Local Similarity 38.2%; Pred. No. 19;  
Matches 13; Conservative 1; Mismatches 9; Indels 11; Gaps 1;

QY 1 VGVSIISAGYQNGFT-----GNITTSAGF 23  
|||||:| |  
Db 69 VGVSIAANISGGHLPVTLGLAVGGNITLTGF 102

RESULT 15  
KICS\_BOVIN  
ID KICS\_BOVIN STANDARD; PRT; 399 AA.  
AC P08728;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-AUG-1988 (Rel. 08, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Keratin, type I cytoskeletal 19 (Cytokeratin 19) (K19) (CK 19).  
GN KRT19.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovioidea;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87004553; PubMed=2428612;





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:46:22 ; Search time 39 Seconds  
(without alignments)  
56.715 Million cell updates/sec

Title: US-09-821-348-2  
Perfect score: 117  
Sequence: 1 VGVSIISAGYQNGFTGNITTSAGF 23

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 76: \*  
1: PIR1: \*  
2: PIR2: \*  
3: PIR3: \*  
4: PIR4: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	101	86.3	875	2 I40862	iota toxin compone
2	54	46.2	551	2 S30691	arylsulfatase (EC
3	54	46.2	551	2 C91220	arylsulfatase [imp
4	54	46.2	551	2 E86066	arylsulfatase [imp
5	53	45.3	749	2 I38488	trophinin - human
6	52	44.4	528	2 AF3125	hypothetical prote
7	52	44.4	538	2 C98162	hypothetical prote
8	51	43.6	3600	2 D86161	Fl003.12 protein -
9	50	42.7	129	2 A12270	hypothetical prote
10	50	42.7	528	2 T41362	hypothetical prote
11	50	42.7	1052	2 AF2959	conserved hypothet
12	50	42.7	1341	2 H98323	hypothetical prote
13	49	41.9	357	2 D82292	phospho-2-dehydro-
14	48	41.0	253	2 A10613	probable lipoprote
15	47	40.2	224	2 G72398	hypothetical prote
16	47	40.2	524	2 T02499	hypothetical prote
17	46	39.3	212	2 E82061	peptide methionine
18	46	39.3	220	2 AD2990	conserved hypothet
19	46	39.3	220	2 E98293	hypothetical prote
20	46	39.3	376	2 AE1878	phospho-2-dehydro-
21	46	39.3	381	2 AG0110	probable exported
22	46	39.3	732	2 F87469	TonB-dependent rec
23	45.5	38.9	250	2 F71442	probable membrane
24	45	38.5	73	2 A27497	Palm I alpha-amyla
25	45	38.5	75	2 A31298	alpha-amylase inhi
26	45	38.5	303	2 AE3520	sugar ABC transpor
27	45	38.5	356	1 ADECHY	2-dehydro-3-deoxy-
28	45	38.5	356	2 H91061	hypothetical prote
29	45	38.5	356	2 D85906	hypothetical prote

30	45	38.5	356	2 AB0833	2-dehydro-3-deoxy-
31	45	38.5	359	1 B69266	acetylpolymine am
32	45	38.5	367	1 I40837	2-dehydro-3-deoxy-
33	45	38.5	565	2 I41061	flagellin - Escher
34	45	38.5	570	2 T07993	ribosomal protein
35	45	38.5	1276	2 B86546	polymorphic outer
36	45	38.5	1276	2 C81591	polymorphic membra
37	45	38.5	1407	2 B72078	polymorphic outer
38	45	38.5	1838	2 T18448	pathogenicity fact
39	44.5	38.0	212	2 F72077	hypothetical prote
40	44.5	38.0	212	2 F86545	hypothetical prote
41	44.5	38.0	250	2 JQ1012	TobRB7-18C protein
42	44.5	38.0	250	2 S13719	probable membrane
43	44.5	38.0	256	2 T48787	hypothetical prote
44	44.5	38.0	312	2 T35413	probable secreted
45	44.5	38.0	399	2 A25470	cytokeratin 19 - b

ALIGNMENTS

RESULT 1

I40862  
iota toxin component Ib - Clostridium perfringens  
C;Species: Clostridium perfringens  
C;Date: 16-Aug-1996 #sequence\_revision 16-Aug-1996 #text\_change 15-Oct-1999  
C;Accession: I40862; S42774  
R;Perelle, S.; Gibert, M.; Boquet, P.; Popoff, M.R.  
Infect. Immun. 61, 5147-5156, 1993  
A;Title: Characterization of Clostridium perfringens iota-toxin genes and expression in  
A;Reference number: I40861; MUID:94041637; PMID:8225592  
A;Accession: I40862  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-875 <RES>  
A;Cross-references: EMBL:X73562; NID:g929031; PIDN:CAA51960.1; PID:g414655

Query Match 86.3%; Score 101; DB 2; Length 875;  
Best Local Similarity 100.0%; Pred. No. 5.8e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VGVSIISAGYQNGFTGNITTS 20

Db 336 VGVSIISAGYQNGFTGNITTS 355

RESULT 2

S30691  
arylsulfatase (EC 3.1.6.1) - Escherichia coli (strain K-12)  
N;Alternate names: asla protein  
C;Species: Escherichia coli  
C;Date: 22-Nov-1993 #sequence\_revision 01-Sep-1995 #text\_change 01-Mar-2002  
C;Accession: S30691; S27555; E65184  
R;Daniels, D.L.; Plunkett III, G.; Burland, V.; Blattner, F.R.  
Science 257, 771-778, 1992  
A;Title: Analysis of the Escherichia coli genome: DNA sequence of the region from 84.5 t  
A;Reference number: S30660; MUID:92358234; PMID:1379743  
A;Accession: S30691  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-551 <DAN>  
A;Cross-references: EMBL:M87049; NID:g836656; PIDN:AAA67597.1; PID:g148200  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1992  
R;Murphy, H.R.; Kalman, M.; Cashel, M.  
submitted to the EMBL Data Library, April 1992  
A;Description: Identification of the gppB locus as two convergent arylsulfatase-like gen  
ia coli.  
A;Reference number: S27554  
A;Accession: S27555  
A;Molecule type: DNA  
A;Residues: 1-190, 'I', 192-461, 'NSLLCVMSSSITS' <MUR>  
A;Cross-references: EMBL:M90498; NID:g145389; PIDN:AAC32036.1; PID:g145391  
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co

A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997

A;Title: The complete genome sequence of *Escherichia coli* K-12.

A;Reference number: A64720; MUID:97426617; PMID:9278503

A;Accession: B65184

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-551 <BLAT>

A;Cross-references: GB:AE000456; GB:U00096; NID:g2367291; PIDN:AAC76804.1; PID:g1790233;

A;Experimental source: strain K-12, substrain MG1655

C;Genetics:

A;Gene: *aslA*; *atsA*

A;Map position: 86 min

C;Function:

A;Description: hydrolyzes phenol sulfates to release sulfate

C;Superfamily: animal sulfatase

C;Keywords: sulfuric ester hydrolase

F;136/Modified site: 3-oxoalanine (Ser) #status predicted

Query Match 46.2%; Score 54; DB 2; Length 551;

Best Local Similarity 56.2%; Pred. No. 3.6;

Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 7 AGYQNGFTGNITTSAG 22

:||| ||||| : |||

Db 486 SGYQGGFTGTVMQTAG 501

#### RESULT 3

C91220

arylsulfatase [imported] - *Escherichia coli* (strain O157:H7, substrain RIMD 0509952)

C;Species: *Escherichia coli*

C;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 17-May-2002

C;Accession: C91220

R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A;Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and gen

A;Reference number: A99629; MUID:21156231; PMID:11258796

A;Accession: C91220

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-551 <HAY>

A;Cross-references: GB:BA000007; PIDN:BA038154.1; PID:g13364207; GSPDB:GN00154

A;Experimental source: strain O157:H7, substrain RIMD 0509952

C;Genetics:

A;Gene: EC84731

C;Superfamily: animal sulfatase

Query Match 46.2%; Score 54; DB 2; Length 551;

Best Local Similarity 56.2%; Pred. No. 3.6;

Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 7 AGYQNGFTGNITTSAG 22

:||| ||||| : |||

Db 486 SGYQGGFTGTVMQTAG 501

#### RESULT 4

E86066

arylsulfatase [imported] - *Escherichia coli* (strain O157:H7, substrain EDL933)

C;Species: *Escherichia coli*

C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 17-May-2002

C;Accession: E86066

R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew

iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,

Nature 409, 529-533, 2001

A;Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.

A;Reference number: A85480; MUID:21074935; PMID:11206551

A;Accession: E86066

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-551 <STO>

A;Cross-references: GB:AE005174; NID:g12518665; PIDN:AAGS8993.1; GSPDB:GN00145; UWGP:Z5;

A;Experimental source: strain O157:H7, substrain EDL933

C;Genetics:

A;Gene: *aslA*

C;Superfamily: animal sulfatase

Query Match 46.2%; Score 54; DB 2; Length 551;

Best Local Similarity 56.2%; Pred. No. 3.6;

Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 7 AGYQNGFTGNITTSAG 22

:||| ||||| : |||

Db 486 SGYQGGFTGTVMQTAG 501

#### RESULT 5

I38488

trophinin - human

C;Species: *Homo sapiens* (man)

C;Date: 16-Feb-1996 #sequence\_revision 16-Feb-1996 #text\_change 21-Jan-2000

C;Accession: I38488

R;Fukuda, M.N.; Sato, T.; Nakayama, J.; Klier, G.; Mikami, M.; Aoki, D.; Nozawa, S.

Genes Dev. 9, 1199-1210, 1995

A;Title: Trophinin and tastin, a novel cell adhesion molecule complex with potential in

A;Reference number: I38487; MUID:95278733; PMID:7758945

A;Accession: I38488

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-749 <RES>

A;Cross-references: EMBL:U04811; NID:g905357; PIDN:AAA79334.1; PID:g836820

C;Superfamily: elastin

Query Match 45.3%; Score 53; DB 2; Length 749;

Best Local Similarity 39.3%; Pred. No. 7;

Matches 11; Conservative 7; Mismatches 4; Indels 6; Gaps 1;

Qy 2 GVSISAGY-----QNGFTGNITTSAGF 23

:||| :||| :||| :||| :|||

Db 530 GLNTSAGFGGGLGTSAGFSGLSTSGF 557

#### RESULT 6

AF3125

hypothetical protein Atu4626 [imported] - *Agrobacterium tumefaciens* (strain C58, Dupont)

C;Species: *Agrobacterium tumefaciens*

C;Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002

C;Accession: AF3125

R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I

erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClell

; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,

ster, E.W.

A;Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.

A;Reference number: AB2577; MUID:21608550; PMID:11743193

A;Accession: AF3125

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-528 <KUR>

A;Cross-references: GB:AE008689; PIDN:AAL45420.1; PID:g17743120; GSPDB:GN00187

A;Experimental source: strain C58 (Dupont)

C;Genetics:

A;Gene: Atu4626

A;Map position: linear chromosome

Query Match 44.4%; Score 52; DB 2; Length 528;

Best Local Similarity 66.7%; Pred. No. 6.7;

Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 7 AGYQNGFTGNITTSAG 21

:||| :||| :||| :|||

Db 356 AGYENGFTLNLTIA 370

RESULT 7  
C98162  
hypothetical protein AGR\_L\_514 [imported] - Agrobacterium tumefaciens (strain C58, Cere  
C;Species: Agrobacterium tumefaciens  
C;Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 18-Nov-2002  
C;Accession: C98162  
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,  
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;  
Science 294, 2323-2328, 2001  
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum  
A;Reference number: A97359; MUID:21608551; PMID:11743194  
A;Accession: C98162  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-538 <KUR>  
A;Cross-references: GB:AE007870; PIDN:AAK88821.1; PID:g15158579; GSPDB:GN00170  
C;Genetics:  
A;Gene: AGR\_L\_514  
A;Map position: linear chromosome

Query Match 44.4%; Score 52; DB 2; Length 538;  
Best Local Similarity 66.7%; Pred. No. 6.9;  
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 7 AGYQNGFTGNITTS A 21  
||||| : : :  
Db 366 AGYENGFTLNKLTIA 380

RESULT 8  
D86161  
F1003.12 protein - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 27-Nov-2001  
C;Accession: D86161  
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A;Reference number: A86141; MUID:21016719; PMID:11130712  
A;Accession: D86161  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-3600 <STO>  
A;Cross-references: GB:AE005172; NID:g4587572; PIDN:AAD25803.1; GSPDB:GN00141  
C;Genetics:  
A;Map position: 1

Query Match 43.6%; Score 51; DB 2; Length 3600;  
Best Local Similarity 60.0%; Pred. No. 75;  
Matches 12; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 3 VSISAGYQNGFTGNITTSAG 22  
||||| : : :  
Db 3422 VPISAIYNDLTGEIVTAAG 3441

RESULT 9  
AI2270  
hypothetical protein alr3720 [imported] - Nostoc sp. (strain PCC 7120)  
C;Species: Nostoc sp. PCC 7120  
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C;Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
C;Accession: AI2270  
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,  
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S

DNA Res. 8, 205-213, 2001  
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana  
A;Reference number: AB1807; MUID:21595285; PMID:11759840  
A;Accession: AI2270  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-129 <KUR>  
A;Cross-references: GB:BA000019; PIDN:BAW75419.1; PID:g17132854; GSPDB:GN00179  
A;Experimental source: strain PCC 7120  
C;Genetics:  
A;Gene: alr3720

Query Match 42.7%; Score 50; DB 2; Length 129;  
Best Local Similarity 45.5%; Pred. No. 2.9;  
Matches 10; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 2 GVSISAGYQNGFTGNITTSAGF 23  
||||| : : :  
Db 94 GVAFSGGVQVGATDNVFGAGY 115

RESULT 10  
T41362

hypothetical protein SPCC4G3.14 - fission yeast (Schizosaccharomyces pombe)  
C;Species: Schizosaccharomyces pombe  
C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 02-Sep-2000  
C;Accession: T41362  
R;Wood, V.; Rajandream, M.A.; Barrell, B.G.; Hilbert, H.; Duesterhoeft, A.  
submitted to the EMBL Data Library, March 1998  
A;Reference number: Z21918  
A;Accession: T41362  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-528 <WOO>  
A;Cross-references: EMBL:Z97052; PIDN:CAB09769.1; GSPDB:GN00068; SPDB:SPCC4G3.14  
A;Experimental source: strain 972h-; cosmid c4G3  
C;Genetics:  
A;Map position: 3  
C;Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology  
F;86-150/Domain: dnaJ amino-terminal homology <DNJ>

Query Match 42.7%; Score 50; DB 2; Length 528;  
Best Local Similarity 56.2%; Pred. No. 13;  
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 8 GYQNGFTGNITTSAGF 23  
||||| : : :  
Db 166 GFQNGFAGASSFSGPF 181

RESULT 11  
AF2959

conserved hypothetical protein Atu3276 [imported] - Agrobacterium tumefaciens (strain C5  
C;Species: Agrobacterium tumefaciens  
C;Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
C;Accession: AF2959  
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L  
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClell  
; Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
ster, E.W.

A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A;Reference number: AB2577; MUID:21608550; PMID:11743193  
A;Accession: AF2959  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1052 <KUR>

A;Cross-references: GB:AE008689; PIDN:AAL44092.1; PID:g17741659; GSPDB:GN00187  
A;Experimental source: strain C58 (Dupont)  
C;Genetics:  
A;Gene: Atu3276



A;Map position: linear chromosome

Query Match 42.7%; Score 50; DB 2; Length 1052;  
Best Local Similarity 60.0%; Pred. No. 28;  
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 8 GYQNGFTGNITTSAG 22  
||:|:||||:|:|  
Db 244 GYASGATGNVTVSNG 258

RESULT 12

H98323  
hypothetical protein AGR\_L\_3085 [imported] - Agrobacterium tumefaciens (strain C58, Cere  
C;Species: Agrobacterium tumefaciens  
C;Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 18-Nov-2002  
C;Accession: H98323  
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman,  
A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.;  
Science 294, 2323-2328, 2001  
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum  
A;Reference number: A97359; MUID:21608551; PMID:11743194  
A;Accession: H98323  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1341 <KUR>  
A;Cross-references: GB:AE007870; PIDN:AAK90114.1; PID:g15160106; GSPDB:GN00170  
C;Genetics:  
A;Gene: AGR\_L\_3085  
A;Map position: linear chromosome

Query Match 42.7%; Score 50; DB 2; Length 1341;  
Best Local Similarity 60.0%; Pred. No. 37;  
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 8 GYQNGFTGNITTSAG 22  
||:|:||||:|:|  
Db 533 GYASGATGNVTVSNG 547

RESULT 13

D82292  
phospho-2-dehydro-3-deoxyheptonate aldolase, tyr-sensitive VC0695 [imported] - Vibrio ch  
C;Species: Vibrio cholerae  
C;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
C;Accession: D82292  
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;  
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, B  
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000  
A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
A;Reference number: A82035; MUID:20406833; PMID:10952301  
A;Accession: D82292  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-357 <HEI>  
A;Cross-references: GB:AE004155; GB:AE003852; NID:g9655127; PIDN:AAF93860.1; GSPDB:GN001  
A;Experimental source: serogroup O1; strain N16961; biotype El Tor  
C;Genetics:  
A;Gene: VC0695  
A;Map position: 1  
C;Superfamily: phospho-2-dehydro-3-deoxyheptonate aldolase

Query Match 41.9%; Score 49; DB 2; Length 357;  
Best Local Similarity 42.1%; Pred. No. 12;  
Matches 8; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 2 GVSISAGYQNGFTGNITTS 20  
|:|:|:|:|:|:|:|:|  
Db 179 GLSMPIGFKNGTDGNLATA 197

RESULT 14

AI0613

probable lipoprotein STY0979 [imported] - Salmonella enterica subsp. enterica serovar Ty  
C;Species: Salmonella enterica subsp. enterica serovar Typhi  
A;Note: this species has also been called Salmonella typhi  
C;Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
C;Accession: AI0613  
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,  
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,  
S.; Moule, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov  
A;Reference number: AB0502; MUID:21534947; PMID:11677608  
A;Accession: AI0613  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-253 <PAR>  
A;Cross-references: GB:AL513382; PIDN:CAD05379.1; PID:g16502142; GSPDB:GN00176  
C;Genetics:  
A;Gene: STY0979

Query Match 41.0%; Score 48; DB 2; Length 253;  
Best Local Similarity 36.4%; Pred. No. 12;  
Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 1 VGVSISAGYQNGFTGNITTSAG 22  
: : :|:|:|:|:|:|:|  
Db 12 ISATLLAGCKNGVNGNLIASSG 33

RESULT 15

G72398  
hypothetical protein TM0246 - Thermotoga maritima (strain MSB8)  
C;Species: Thermotoga maritima  
C;Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C;Accession: G72398  
R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;  
C.M.

Nature 399, 323-329, 1999  
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq  
A;Reference number: A72200; MUID:99287316; PMID:10360571  
A;Accession: G72398  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-224 <ARN>  
A;Cross-references: GB:AE001708; GB:AE000512; NID:g4980740; PIDN:AAD35337.1; PID:g498074  
A;Experimental source: strain MSB8  
C;Genetics:  
A;Gene: TM0246  
C;Superfamily: Thermotoga maritima hypothetical protein TM0246

Query Match 40.2%; Score 47; DB 2; Length 224;  
Best Local Similarity 47.4%; Pred. No. 15;  
Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 5 ISAGYQNGFTGNITTSAGF 23  
: :|:|:|:|:|:|:|  
Db 103 VLSGYAPGFGGNVTVVACF 121

Search completed: July 30, 2003, 16:51:11  
Job time : 42 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:37:37 ; Search time 51 Seconds  
(without alignments)  
53.558 Million cell updates/sec

Title: US-09-821-348-2  
Perfect score: 117  
Sequence: 1 VGVSTISAGYQNGFTGNITTSAGF 23

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09D\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09E\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	51	43.6	557	9	US-09-864-761-34705
2	48	41.0	831	9	US-09-789-561-86
3	47	40.2	262	15	US-10-156-761-14006
4	46	39.3	375	15	US-10-156-761-8098
5	45	38.5	366	10	US-09-738-626-4603
6	45	38.5	871	9	US-09-886-468-21
7	45	38.5	1838	9	US-09-879-248-8
8	44.5	38.0	943	9	US-09-815-242-11834
9	44	37.6	128	10	US-09-925-300-1730
10	44	37.6	400	9	US-09-922-217-1115
11	44	37.6	400	14	US-10-025-380-1115
12	44	37.6	424	10	US-09-738-626-4306
13	44	37.6	476	8	US-08-834-705-4
14	44	37.6	889	11	US-09-952-267-15
15	43	36.8	191	10	US-09-738-626-4151

16	43	36.8	212	10	US-09-738-626-6254	Sequence 6254, Ap
17	43	36.8	367	10	US-09-881-752A-188	Sequence 188, App
18	43	36.8	369	11	US-09-988-067B-22	Sequence 22, Appl
19	43	36.8	409	15	US-10-013-315-24	Sequence 24, Appl
20	43	36.8	503	9	US-09-737-178-52	Sequence 52, Appl
21	43	36.8	503	10	US-09-286-488-52	Sequence 52, Appl
22	43	36.8	503	11	US-09-853-079-52	Sequence 52, Appl
23	43	36.8	549	9	US-09-815-242-11313	Sequence 11313, A
24	43	36.8	666	9	US-09-737-178-85	Sequence 85, Appl
25	43	36.8	666	11	US-09-853-079-85	Sequence 85, Appl
26	43	36.8	677	9	US-09-737-178-144	Sequence 144, App
27	43	36.8	677	11	US-09-853-079-144	Sequence 144, App
28	43	36.8	1132	9	US-09-737-178-87	Sequence 87, Appl
29	43	36.8	1132	11	US-09-853-079-87	Sequence 87, Appl
30	42.5	36.3	388	10	US-09-764-864-931	Sequence 931, App
31	42	35.9	118	9	US-09-925-301-1454	Sequence 1454, Ap
32	42	35.9	281	13	US-10-010-901-40	Sequence 40, Appl
33	42	35.9	314	15	US-10-200-910-4	Sequence 4, Appli
34	42	35.9	314	15	US-10-200-910-12	Sequence 12, Appl
35	42	35.9	351	15	US-10-200-910-2	Sequence 2, Appli
36	42	35.9	351	15	US-10-200-910-10	Sequence 10, Appl
37	42	35.9	399	15	US-10-200-910-8	Sequence 8, Appli
38	42	35.9	436	14	US-10-076-535-2	Sequence 2, Appli
39	42	35.9	436	15	US-10-200-910-6	Sequence 6, Appli
40	42	35.9	492	9	US-09-737-178-21	Sequence 21, Appl
41	42	35.9	492	10	US-09-286-488-21	Sequence 21, Appl
42	42	35.9	492	11	US-09-853-079-21	Sequence 21, Appl
43	42	35.9	826	9	US-09-741-669-309	Sequence 309, App
44	42	35.9	826	9	US-09-912-020-286	Sequence 286, App
45	42	35.9	833	15	US-10-156-761-12366	Sequence 12366, A

ALIGNMENTS

RESULT 1

US-09-864-761-34705  
; Sequence 34705, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Aeomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34705
; LENGTH: 557
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL049732.9
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.8
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.3
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3.9
; OTHER INFORMATION: SWISSPROT HIT: Q12816, EVALUE 0.00e+00
; OTHER INFORMATION: EST\_HUMAN HIT: BE894856.1, EVALUE 1.00e-106
US-09-864-761-34705

Query Match 43.6%; Score 51; DB 9; Length 557;
Best Local Similarity 45.0%; Pred. No. 11;
Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

Qy 4 SISAGYQNGFTGNITTSAGF 23
: | | : | | : | | : | |
Db 214 TFSGGASSGFGGTLSTTAGF 233

RESULT 2
US-09-789-561-86
; Sequence 86, Application US/09789561
; Patent No. US20020064818A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 52 Human secreted proteins
; FILE REFERENCE: PZ043P1
; CURRENT APPLICATION NUMBER: US/09/789,561
; CURRENT FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: PCT/US00/24008
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 60/152,317
; PRIOR FILING DATE: 1999-09-03
; PRIOR APPLICATION NUMBER: 60/152,315
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 86
; LENGTH: 831
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-789-561-86

Query Match 41.0%; Score 48; DB 9; Length 831;
Best Local Similarity 47.4%; Pred. No. 50;
Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Qy 2 GVSISAGYQNGFTGNITTS 20
| : | | : | | : | | : | |

Db 685 GLAIGAAFSAGLTGGISTS 703
RESULT 3
US-10-156-761-14006
; Sequence 14006, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 14006
; LENGTH: 262
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-14006

Query Match 40.2%; Score 47; DB 15; Length 262;
Best Local Similarity 45.0%; Pred. No. 19;
Matches 9; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

Qy 1 VGVSIISAGYQNGFTGNITTS 20
| : | | : | | : | | : | |
Db 207 VGLLISVAYENGHGGAVSAS 226

RESULT 4
US-10-156-761-8098
; Sequence 8098, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 8098
; LENGTH: 375
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-8098

Query Match 39.3%; Score 46; DB 15; Length 375;
Best Local Similarity 36.4%; Pred. No. 41;
Matches 8; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

Qy 2 GVSISAGYQNGFTGNITTSAGF 23
| : | | : | | : | | : | |
Db 344 GGRVSFGQSGYSGTFAAPAGF 365

RESULT 5  
US-09-738-626-4603  
; Sequence 4603, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 4603  
; LENGTH: 366  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-4603

Query Match 38.5%; Score 45; DB 10; Length 366;  
Best Local Similarity 50.0%; Pred. No. 56;  
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 2 GVSISAGYQNGFTGNI 17  
|:|: |:|: |||  
Db 186 GMSMPIGPKNGTDGNI 201

RESULT 6  
US-09-886-468-21  
; Sequence 21, Application US/09886468  
; Patent No. US2002037293A1  
; GENERAL INFORMATION:  
; APPLICANT: Aventis Pasteur Limited  
; TITLE OF INVENTION: Chlamydia antigens and corresponding DNA fragments and uses there  
; FILE REFERENCE: 77813-5  
; CURRENT APPLICATION NUMBER: US/09/886,468  
; CURRENT FILING DATE: 1999-12-23  
; PRIOR APPLICATION NUMBER: 60/113,280  
; PRIOR FILING DATE: 1998-12-23  
; PRIOR APPLICATION NUMBER: 60/113,281  
; PRIOR FILING DATE: 1998-12-23  
; PRIOR APPLICATION NUMBER: 60/113,282  
; PRIOR FILING DATE: 1998-12-23  
; PRIOR APPLICATION NUMBER: 60/113,283  
; PRIOR FILING DATE: 1998-12-23  
; PRIOR APPLICATION NUMBER: 60/113,284  
; PRIOR FILING DATE: 1998-12-23  
; PRIOR APPLICATION NUMBER: 60/113,285  
; PRIOR FILING DATE: 1998-12-23  
; PRIOR APPLICATION NUMBER: 60/113,385  
; PRIOR FILING DATE: 1998-12-23  
; PRIOR APPLICATION NUMBER: 60/114,050  
; PRIOR FILING DATE: 1998-12-28  
; PRIOR APPLICATION NUMBER: 60/114,056  
; PRIOR FILING DATE: 1998-12-28  
; PRIOR APPLICATION NUMBER: 60/114,057

; PRIOR FILING DATE: 1998-12-28  
; PRIOR APPLICATION NUMBER: 60/114,058  
; PRIOR FILING DATE: 1998-12-28  
; PRIOR APPLICATION NUMBER: 60/114,059  
; PRIOR FILING DATE: 1998-12-28  
; PRIOR APPLICATION NUMBER: 60/114,061  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 21  
; LENGTH: 871  
; TYPE: PRT  
; ORGANISM: Chlamydia pneumoniae  
US-09-886-468-21

Query Match 38.5%; Score 45; DB 9; Length 871;  
Best Local Similarity 42.9%; Pred. No. 1.5e+02;  
Matches 9; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

Qy 2 GVSISAGYQNGFTGNTTSAG 22  
|:|: |:|: |||: |  
Db 309 GLAISQNMFSNTTANG 329

RESULT 7  
US-09-879-248-8  
; Sequence 8, Application US/09879248  
; Patent No. US20020062500A1  
; GENERAL INFORMATION:  
; APPLICANT: Fan, Hao  
; APPLICANT: Wei, Zhong-Min  
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITING DOMAINS AND USE  
; TITLE OF INVENTION: THEREOF  
; FILE REFERENCE: 21829/81  
; CURRENT APPLICATION NUMBER: US/09/879,248  
; CURRENT FILING DATE: 2001-06-12  
; PRIOR APPLICATION NUMBER: 60/212,211  
; PRIOR FILING DATE: 2000-06-16  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 8  
; LENGTH: 1838  
; TYPE: PRT  
; ORGANISM: Erwinia amylovora  
US-09-879-248-8

Query Match 38.5%; Score 45; DB 9; Length 1838;  
Best Local Similarity 38.1%; Pred. No. 3.5e+02;  
Matches 8; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

Qy 2 GVSISAGYQNGFTGNTTSAG 22  
|:|: |:|: |:|: |  
Db 1329 GLNVSGRDGGVSGNIMVATG 1349

RESULT 8  
US-09-815-242-11834  
; Sequence 11834, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11834
; LENGTH: 943
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-11834

Query Match 38.0%; Score 44.5; DB 9; Length 943;
Best Local Similarity 39.1%; Pred. No. 2e+02;
Matches 9; Conservative 4; Mismatches 7; Indels 3; Gaps 1;

QY 3 VSISAG---YQNGFTGNITTSAG 22
| : | | | | | : | : | : |
Db 294 VELGSGDVKYHQGFSSNVMTSGG 316

RESULT 9
US-09-925-300-1730
; Sequence 1730, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1730
; LENGTH: 128
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-300-1730

Query Match 37.6%; Score 44; DB 10; Length 128;
Best Local Similarity 45.5%; Pred. No. 24;
Matches 10; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 1 VGVSIISAGYQNGFTGNITTSAG 22
| | | | | | : | : | : |
Db 70 VSSSSSGGYGGYGGVLTASDG 91

RESULT 10
US-09-922-217-1115
; Sequence 1115, Application US/09922217
; Patent No. US20020076414A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy

; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Smith, Carole Lynn
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.471C13
; CURRENT APPLICATION NUMBER: US/09/922,217
; CURRENT FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1115
; LENGTH: 400
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-217-1115

Query Match 37.6%; Score 44; DB 9; Length 400;
Best Local Similarity 45.5%; Pred. No. 88;
Matches 10; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 1 VGVSIISAGYQNGFTGNITTSAG 22
| | | | | | : | : | : |
Db 53 VSSSSSGGYGGYGGVLTASDG 74

RESULT 11
US-10-025-380-1115
; Sequence 1115, Application US/10025380
; Publication No. US20020182191A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Smith, Carole L.
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Skeiky, Yasir A. W.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick Thomas S.
; APPLICANT: Carter, Darrick
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.471C14
; CURRENT APPLICATION NUMBER: US/10/025,380
; CURRENT FILING DATE: 2001-12-19
; NUMBER OF SEQ ID NOS: 1129
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1115
; LENGTH: 400
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-025-380-1115

Query Match 37.6%; Score 44; DB 14; Length 400;
Best Local Similarity 45.5%; Pred. No. 88;
Matches 10; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 1 VGVSIISAGYQNGFTGNITTSAG 22
| | | | | | : | : | : |
Db 53 VSSSSSGGYGGYGGVLTASDG 74

RESULT 12

US-09-738-626-4306  
; Sequence 4306, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 4306  
; LENGTH: 424  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-4306  
  
Query Match 37.6%; Score 44; DB 10; Length 424;  
Best Local Similarity 47.6%; Pred. No. 94;  
Matches 10; Conservative 0; Mismatches 11; Indels 0; Gaps 0;  
  
Qy 1 VGVSTLGGYNGINVNSENKA 21  
Db 305 VGVSTLGGYNGINVNSENKA 325  
  
RESULT 13  
US-08-834-705-4  
; Sequence 4, Application US/08834705  
; Publication No. US20030023066A1  
; GENERAL INFORMATION:  
; APPLICANT: Haas, Rainer et al.  
; TITLE OF INVENTION: New Helicobacter Polypeptides  
; TITLE OF INVENTION: and Corresponding Polynucleotide Molecules  
; NUMBER OF SEQUENCES: 66  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Clark & Elbing LLP  
; STREET: 176 Federal Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/834,705  
; FILING DATE: 01-APR-1997  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/749,051  
; FILING DATE: 14-NOV-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Clark, Paul T.  
; REGISTRATION NUMBER: 30,175  
; REFERENCE/DOCKET NUMBER: 06132/039001

TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-428-0200  
; TELEFAX: 617-428-7045  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 476 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: internal  
; FEATURE:  
; NAME/KEY: Signal Sequence  
; LOCATION: 1..25  
; OTHER INFORMATION:  
US-08-834-705-4  
  
Query Match 37.6%; Score 44; DB 8; Length 476;  
Best Local Similarity 45.0%; Pred. No. 1.1e+02;  
Matches 9; Conservative 2; Mismatches 9; Indels 0; Gaps 0;  
  
Qy 1 VGVSTLGGYNGINVNSENKA 21  
Db 305 VGVSTLGGYNGINVNSENKA 325  
  
RESULT 14  
US-09-952-267-15  
; Sequence 15, Application US/09952267  
; Publication No. US20030032772A1  
; GENERAL INFORMATION:  
; APPLICANT: HANSEN, ERIC J.  
; APPLICANT: AEBI, CHRISTOPH  
; APPLICANT: COPE, LESLIE D.  
; APPLICANT: MACIVER, ISOBEL  
; APPLICANT: FISKE, MICHAEL J.  
; APPLICANT: FREDENBURG, ROSS A.  
; TITLE OF INVENTION: USP1 AND USP2 ANTIGENS OF MORAXELLA CATARRHALIS  
; FILE REFERENCE: AMCY:024  
; CURRENT APPLICATION NUMBER: US/09/952,267  
; CURRENT FILING DATE: 2001-09-12  
; PRIOR APPLICATION NUMBER: 09/336,447  
; PRIOR FILING DATE: 1999-06-21  
; NUMBER OF SEQ ID NOS: 98  
; SOFTWARE: PatentIn ver. 2.1  
; SEQ ID NO 15  
; LENGTH: 889  
; TYPE: PRT  
; ORGANISM: Moraxella catarrhalis  
US-09-952-267-15  
  
Query Match 37.6%; Score 44; DB 11; Length 889;  
Best Local Similarity 50.0%; Pred. No. 2.2e+02;  
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;  
  
Qy 5 ISAGYQNGFTGNITTSAG 22  
Db 175 VSGGYANQATGESSTVAG 192  
  
RESULT 15  
US-09-738-626-4151  
; Sequence 4151, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAKO

Search completed: July 30, 2003, 16:46:18  
Job time : 52 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:39:04 ; Search time 17 Seconds  
(without alignments)  
57.244 Million cell updates/sec

Title: US-09-821-348-2  
Perfect score: 117  
Sequence: 1 VGVISAGYQNGFTGNITTSAGF 23

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	55	47.0	674	1 US-08-317-522A-3	Sequence 3, Appli
2	55	47.0	674	1 US-08-439-818A-3	Sequence 3, Appli
3	55	47.0	674	2 US-08-751-965-3	Sequence 3, Appli
4	55	47.0	674	2 US-08-738-975-3	Sequence 3, Appli
5	55	47.0	674	2 US-08-728-626-3	Sequence 3, Appli
6	55	47.0	674	3 US-08-808-599A-3	Sequence 3, Appli
7	53	45.3	749	1 US-08-317-522A-2	Sequence 2, Appli
8	53	45.3	749	1 US-08-439-818A-2	Sequence 2, Appli
9	53	45.3	749	2 US-08-751-965-2	Sequence 2, Appli
10	53	45.3	749	2 US-08-738-975-2	Sequence 2, Appli
11	53	45.3	749	2 US-08-728-626-2	Sequence 2, Appli
12	53	45.3	749	3 US-08-808-599A-2	Sequence 2, Appli
13	51	43.6	150	3 US-08-808-599A-39	Sequence 39, Appli
14	51	43.6	1160	3 US-08-808-599A-24	Sequence 24, Appli
15	47	40.2	82	3 US-08-808-599A-38	Sequence 38, Appli
16	47	40.2	86	1 US-08-439-818A-22	Sequence 22, Appli
17	47	40.2	86	2 US-08-751-965-22	Sequence 22, Appli
18	47	40.2	86	2 US-08-738-975-22	Sequence 22, Appli
19	47	40.2	86	2 US-08-728-626-22	Sequence 22, Appli
20	47	40.2	86	3 US-08-808-599A-22	Sequence 22, Appli
21	45	38.5	1132	4 US-09-198-452A-466	Sequence 466, App
22	45	38.5	1838	3 US-09-120-663-2	Sequence 2, Appli
23	44.5	38.0	220	4 US-09-198-452A-461	Sequence 461, App
24	44.5	38.0	250	1 US-08-234-939-2	Sequence 2, Appli
25	44.5	38.0	250	1 US-08-558-865-2	Sequence 2, Appli
26	44.5	38.0	250	3 US-08-654-025-2	Sequence 2, Appli
27	44.5	38.0	250	3 US-08-654-025-7	Sequence 7, Appli

28	44.5	38.0	454	4 US-09-252-991A-25978	Sequence 25978, A
29	44	37.6	362	4 US-09-252-991A-30517	Sequence 30517, A
30	44	37.6	400	1 US-07-730-953-2	Sequence 2, Appli
31	44	37.6	513	1 US-08-200-232-4	Sequence 4, Appli
32	44	37.6	513	5 PCT-US95-02219-4	Sequence 4, Appli
33	44	37.6	513	5 PCT-US95-02219A-4	Sequence 4, Appli
34	44	37.6	889	4 US-09-336-447A-15	Sequence 15, Appli
35	44	37.6	986	3 US-08-938-291A-7	Sequence 7, Appli
36	44	37.6	986	4 US-09-589-619-7	Sequence 7, Appli
37	43	36.8	82	3 US-08-808-599A-37	Sequence 37, Appli
38	43	36.8	503	3 US-08-845-258-52	Sequence 52, Appli
39	43	36.8	503	3 US-08-990-571-52	Sequence 52, Appli
40	43	36.8	503	4 US-09-528-784A-52	Sequence 52, Appli
41	43	36.8	503	4 US-09-569-098A-52	Sequence 52, Appli
42	43	36.8	666	4 US-09-528-784A-85	Sequence 85, Appli
43	43	36.8	666	4 US-09-569-098A-85	Sequence 85, Appli
44	43	36.8	792	3 US-08-433-522A-8	Sequence 8, Appli
45	43	36.8	792	3 US-09-135-166-8	Sequence 8, Appli

ALIGNMENTS

RESULT 1  
US-08-317-522A-3  
; Sequence 3, Application US/08317522A  
; Patent No. 559918  
; GENERAL INFORMATION:  
; APPLICANT: Fukuda, Michiko N.  
; TITLE OF INVENTION: Trophinin and Trophinin-Assisting  
; TITLE OF INVENTION: Proteins  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell and Flores  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/317,522A  
; FILING DATE: 04-OCT-1994  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn A.  
; REGISTRATION NUMBER: 31,815  
; REFERENCE/DOCKET NUMBER: P-LA 9991  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 674 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
US-08-317-522A-3

Query Match 47.0%; Score 55; DB 1; Length 674;  
Best Local Similarity 50.0%; Pred. No. 3;  
Matches 11; Conservative 1; Mismatches 10; Indels 0; Gaps 0;

Qy 2 GVSISAGYQNGFTGNITTSAGF 23  
Db 446 GFSFGNGLSTGFGGLNTSAGF 467

RESULT 2  
US-08-439-818A-3



; Sequence 3, Application US/08439818A  
; Patent No. 5654145  
; GENERAL INFORMATION:  
; APPLICANT: Fukuda, Michiko N.  
; TITLE OF INVENTION: Trophinin and Trophinin-Assisting  
; TITLE OF INVENTION: Proteins  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell and Flores  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/439,818A  
; FILING DATE: 12-MAY-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/317,522  
; FILING DATE: 04-OCT-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn A.  
; REGISTRATION NUMBER: 31,815  
; REFERENCE/DOCKET NUMBER: P-LA 1563  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 674 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; US-08-439-818A-3

Query Match 47.0%; Score 55; DB 1; Length 674;  
Best Local Similarity 50.0%; Pred. No. 3;  
Matches 11; Conservative 1; Mismatches 10; Indels 0; Gaps 0;

QY 2 GVSISAGYQNGFTGNITTSAGF 23  
Db 446 GFSFGNGLSTGFGGLNTSAGF 467

RESULT 3  
US-08-751-965-3  
; Sequence 3, Application US/08751965  
; Patent No. 5858360  
; GENERAL INFORMATION:  
; APPLICANT: Fukuda, Michiko N.  
; TITLE OF INVENTION: Trophinin and Trophinin-Assisting  
; TITLE OF INVENTION: Proteins  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell and Flores  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/751,965  
; FILING DATE: Herewith

; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/439,818  
; FILING DATE: 12-MAY-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn A.  
; REGISTRATION NUMBER: 31,815  
; REFERENCE/DOCKET NUMBER: P-LA 2252  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 674 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; US-08-751-965-3

Query Match 47.0%; Score 55; DB 2; Length 674;  
Best Local Similarity 50.0%; Pred. No. 3;  
Matches 11; Conservative 1; Mismatches 10; Indels 0; Gaps 0;

QY 2 GVSISAGYQNGFTGNITTSAGF 23  
Db 446 GFSFGNGLSTGFGGLNTSAGF 467

RESULT 4  
US-08-738-975-3  
; Sequence 3, Application US/08738975  
; Patent No. 5880267  
; GENERAL INFORMATION:  
; APPLICANT: Fukuda, Michiko N.  
; TITLE OF INVENTION: Trophinin and Trophinin-Assisting  
; TITLE OF INVENTION: Proteins  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell and Flores  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/738,975  
; FILING DATE: herewith  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/439,818  
; FILING DATE: 05-Dec-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn A.  
; REGISTRATION NUMBER: 31,815  
; REFERENCE/DOCKET NUMBER: P-LA 2251  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 674 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; US-08-738-975-3

Query Match 47.0%; Score 55; DB 2; Length 674;  
Best Local Similarity 50.0%; Pred. No. 3;  
Matches 11; Conservative 1; Mismatches 10; Indels 0; Gaps 0;



QY 2 GVSISAGYQNGFTGNITTSAGF 23  
| | | | | : | | | | |  
Db 446 GFSFGNGLSTGFGGLNTSAGF 467

## RESULT 5

US-08-728-626-3  
; Sequence 3, Application US/08728626  
; Patent No. 5910451  
; GENERAL INFORMATION:  
; APPLICANT: Fukuda, Michiko N.  
; TITLE OF INVENTION: Trophinin and Trophinin-Assisting  
; TITLE OF INVENTION: Proteins  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell and Flores  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/728,626  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/439,818  
; FILING DATE: 12-MAY-1995  
; APPLICATION NUMBER: US 08/317,522  
; FILING DATE: 04-OCT-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn A.  
; REGISTRATION NUMBER: 31,815  
; REFERENCE/DOCKET NUMBER: P-LA 1563  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 674 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
US-08-728-626-3

Query Match 47.0%; Score 55; DB 2; Length 674;  
Best Local Similarity 50.0%; Pred. No. 3;  
Matches 11; Conservative 1; Mismatches 10; Indels 0; Gaps 0;

QY 2 GVSISAGYQNGFTGNITTSAGF 23  
| | | | | : | | | | |  
Db 446 GFSFGNGLSTGFGGLNTSAGF 467

## RESULT 6

US-08-808-599A-3  
; Sequence 3, Application US/08808599A  
; Patent No. 6111089  
; GENERAL INFORMATION:  
; APPLICANT: Fukuda, Michiko N.  
; TITLE OF INVENTION: Trophinin, Trophinin-Assisting  
; TITLE OF INVENTION: Proteins and Methods to Inhibit Implantation  
; NUMBER OF SEQUENCES: 41  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell & Flores LLP  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA

ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/808,599A  
; FILING DATE: 28-FEB-1997  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/317,522  
; FILING DATE: 04-OCT-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/439,818  
; FILING DATE: 12-MAY-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn A.  
; REGISTRATION NUMBER: 31,815  
; REFERENCE/DOCKET NUMBER: P-LA 2256  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 674 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
US-08-808-599A-3

Query Match 47.0%; Score 55; DB 3; Length 674;  
Best Local Similarity 50.0%; Pred. No. 3;  
Matches 11; Conservative 1; Mismatches 10; Indels 0; Gaps 0;

QY 2 GVSISAGYQNGFTGNITTSAGF 23  
| | | | | : | | | | |  
Db 446 GFSFGNGLSTGFGGLNTSAGF 467

## RESULT 7

US-08-317-522A-2  
; Sequence 2, Application US/08317522A  
; Patent No. 5599918  
; GENERAL INFORMATION:  
; APPLICANT: Fukuda, Michiko N.  
; TITLE OF INVENTION: Trophinin and Trophinin-Assisting  
; TITLE OF INVENTION: Proteins  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell and Flores  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/317,522A  
; FILING DATE: 04-OCT-1994  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn A.  
; REGISTRATION NUMBER: 31,815  
; REFERENCE/DOCKET NUMBER: P-LA 9991  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:

```
; LENGTH: 749 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-317-522A-2
```

Query Match  
Best Local Similarity    45.3%; Score 53; DB 1; Length 749;  
Matches    11; Conservative    7; Mismatches    4; Indels    6; Gaps    1;

QY         2 GVSIAGY-----QGFTGNITTSAGF 23  
            |::|||:  
Db         530 GLNTSAGFGGLGTSAGFSGGLSTSSGF 557

RESULT 8  
US-08-439-818A-2  
; Sequence 2, Application US/08439818A  
; Patent No. 5654145  
; GENERAL INFORMATION:  
; APPLICANT: Fukuda, Michiko N.  
; TITLE OF INVENTION: Trophinin and Trophinin-Assisting  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell and Flores  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/439,818  
; FILING DATE: Herewith  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/439,818  
; FILING DATE: 12-MAY-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn A.  
; REGISTRATION NUMBER: 31,815  
; REFERENCE/DOCKET NUMBER: P-LA 2252  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 749 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-751-965-2

Query Match          45.3%; Score 53; DB 2; Length  
Best Local Similarity    39.3%; Pred. No. 6.7;  
Matches    11; Conservative    7; Mismatches    4; Inde

QY         2 GVSISAGY-----QNFTGNITTSAGF 23  
            ||::|||:  
Db         530 GLNLSAGFGGLGTSGAGFSGLTSSGF 557

RESULT 10  
US-08-738-975-2  
; Sequence 2, Application US/08738975  
; Patent No. 5880267  
; GENERAL INFORMATION:  
; APPLICANT: Fukuda, Michiko N.  
; TITLE OF INVENTION: Trophinin and Trophinin-Assisting  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell and Flores  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/738,975  
; FILING DATE: herewith  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/738,975  
; FILING DATE: herewith  
; CLASIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/439,818

;; FILING DATE: 05-Dec-1995  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Campbell, Cathryn A.  
;; REGISTRATION NUMBER: 31,815  
;; REFERENCE/DOCKET NUMBER: P-LA 2251  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (619) 535-9001  
;; TELEFAX: (619) 535-8949  
;; INFORMATION FOR SEQ ID NO: 2:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 749 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-08-738-975-2

Query Match 45.3%; Score 53; DB 2; Length 749;  
Best Local Similarity 39.3%; Pred. No. 6.7;  
Matches 11; Conservative 7; Mismatches 4; Indels 6; Gaps 1;

QY 2 GVSISAGY-----QNGFTGNITTSAGF 23  
|::|||: ||:|::||:|  
Db 530 GLNTSAGFGGGLGTSAGFSGGLSTSSGF 557

RESULT 11  
US-08-728-626-2  
; Sequence 2, Application US/08728626  
; Patent No. 5910451  
; GENERAL INFORMATION:  
; APPLICANT: Fukuda, Michiko N.  
; TITLE OF INVENTION: Trophinin and Trophinin-Assisting  
; TITLE OF INVENTION: Proteins  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell and Flores  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/728,626  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/439,818  
; FILING DATE: 12-MAY-1995  
; APPLICATION NUMBER: US 08/317,522  
; FILING DATE: 04-OCT-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn A.  
; REGISTRATION NUMBER: 31,815  
; REFERENCE/DOCKET NUMBER: P-LA 1563  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 749 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-728-626-2

Query Match 45.3%; Score 53; DB 2; Length 749;  
Best Local Similarity 39.3%; Pred. No. 6.7;  
Matches 11; Conservative 7; Mismatches 4; Indels 6; Gaps 1;

QY 2 GVSISAGY-----QNGFTGNITTSAGF 23  
|::|||: ||:|::||:|  
Db 530 GLNTSAGFGGGLGTSAGFSGGLSTSSGF 557  
  
RESULT 12  
US-08-808-599A-2  
; Sequence 2, Application US/08808599A  
; Patent No. 6111089  
; GENERAL INFORMATION:  
; APPLICANT: Fukuda, Michiko N.  
; TITLE OF INVENTION: Trophinin, Trophinin-Assisting  
; TITLE OF INVENTION: Proteins and Methods to Inhibit Implantation  
; NUMBER OF SEQUENCES: 41  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell & Flores LLP  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/808,599A  
; FILING DATE: 28-FEB-1997  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/317,522  
; FILING DATE: 04-OCT-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/439,818  
; FILING DATE: 12-MAY-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn A.  
; REGISTRATION NUMBER: 31,815  
; REFERENCE/DOCKET NUMBER: P-LA 2256  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 749 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-808-599A-2

Query Match 45.3%; Score 53; DB 3; Length 749;  
Best Local Similarity 39.3%; Pred. No. 6.7;  
Matches 11; Conservative 7; Mismatches 4; Indels 6; Gaps 1;

QY 2 GVSISAGY-----QNGFTGNITTSAGF 23  
|::|||: ||:|::||:|  
Db 530 GLNTSAGFGGGLGTSAGFSGGLSTSSGF 557

RESULT 13  
US-08-808-599A-39  
; Sequence 39, Application US/08808599A  
; Patent No. 6111089  
; GENERAL INFORMATION:  
; APPLICANT: Fukuda, Michiko N.  
; TITLE OF INVENTION: Trophinin, Trophinin-Assisting  
; TITLE OF INVENTION: Proteins and Methods to Inhibit Implantation  
; NUMBER OF SEQUENCES: 41  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell & Flores LLP  
; STREET: 4370 La Jolla Village Drive, Suite 700

CITY: San Diego  
STATE: California  
COUNTRY: USA  
ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/808,599A  
FILING DATE: 28-FEB-1997  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/317,522  
FILING DATE: 04-OCT-1994  
APPLICATION DATA:  
APPLICATION NUMBER: US 08/439,818  
FILING DATE: 12-MAY-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-LA 2256  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 39:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 150 amino acids  
TYPE: amino acid  
TOPOLOGY: linear

```
Query Match      43.6%; Score 51; DB 3; Length 150;
Best Local Similarity 45.0%; Pred. No. 2.1;
Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;
```

QY	4	SISAGYQNGFTGNITTSAGF	23
		:     :     :	
Db	82	TFSGCASSGFGCTLSTTAGF	101

RESULT 14  
US-08-808-599A-24  
; Sequence 24, Application US/08808599A  
; Patent No. 6111089  
; GENERAL INFORMATION:  
; APPLICANT: Fukuda, Michiko N.  
; TITLE OF INVENTION: Trophinin, Trophinin-Assisting  
; TITLE OF INVENTION: Proteins and Methods to Inhibit Implantation  
; NUMBER OF SEQUENCES: 41  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell & Flores LLP  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/808,599A  
; FILING DATE: 28-FEB-1997  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/317,522  
; FILING DATE: 04-OCT-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/439,818  
; FILING DATE: 12-MAY-1995

ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-LA 2256  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1160 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-808-599A-24

Query Match 43.6%; Score 51; DB 3; Length 1160;  
Best Local Similarity 42.9%; Pred. No. 22;  
Matches 12; Conservative 4; Mismatches 6; Indels 6; Gaps 1;

QY 2 GVSISAGYQN-----GFTGNITTSAGF 23  
| : | | | : | | | : | |  
Db 827 GLTSAGFGNGLGTSAAGFDSSLGTSTGF 854

RESULT 15  
US-08-808-599A-38  
; Sequence 38, Application US/08808599A  
; Patent No. 6111089  
; GENERAL INFORMATION:  
; APPLICANT: Fukuda, Michiko N.  
; TITLE OF INVENTION: Trophinin, Trophinin-Assisting  
; TITLE OF INVENTION: Proteins and Methods to Inhibit Implantation  
; NUMBER OF SEQUENCES: 41  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell & Flores LLP  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92122

Query Match 40.2%; Score 47; DB 3; Length 82;  
Best Local Similarity 39.1%; Pred. No. 4.2;  
Matches 9; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

Oy 1 VGVSISAGYQNGFTGNITTSAGF 23  
:| :||| :|||  
Db 19 IGFGSGSNTSTGFTGEPSTSTGF 41

Search completed: July 30, 2003, 16:48:41  
Job time : 18 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run On: July 30, 2003, 16:38:07 ; Search time 83 seconds  
(without alignments)  
43.984 Million cell updates/sec

Title: US-09-821-348-2  
Perfect score: 117  
Sequence: 1 VGVSIAGYQNGFTGNTTSAGF 23

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_19Jun03:\*  
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*  
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	55	47.0	674	21 AAB23358	Human trophinin pr
2	54	46.2	305	22 ABG13155	Novel human diagno
3	54	46.2	306	22 ABG08885	Novel human diagno
4	54	46.2	314	22 ABG13158	Novel human diagno
5	54	46.2	478	22 ABG08884	Novel human diagno
6	54	46.2	478	22 ABG13156	Novel human diagno
7	54	46.2	822	22 ABG08883	Novel human diagno
8	54	46.2	822	22 ABG24362	Novel human diagno
9	54	46.2	822	22 ABG29163	Novel human diagno

10	54	46.2	861	22	ABG18082	Novel human diagno
11	54	46.2	861	22	ABG22537	Novel human diagno
12	53	45.3	749	17	AAR94895	Human trophinin.
13	53	45.3	749	21	AAB23357	Human trophinin pr
14	53	45.3	814	22	AAM94034	Human stomach canc
15	53	45.3	814	22	AAB94744	Human protein sequ
16	53	45.3	1406	22	AAU32728	Novel human secret
17	51	43.6	150	21	AAB23378	Human trophinin co
18	51	43.6	557	22	ABG48792	Human liver peptid
19	51	43.6	557	22	ABB28781	Peptide #1432 enco
20	51	43.6	557	22	ABB33968	Peptide #1474 enco
21	51	43.6	557	22	ABB19407	Protein #1406 enco
22	51	43.6	557	22	AAM54732	Human brain expres
23	51	43.6	557	22	AAM67128	Human bone marrow
24	51	43.6	557	22	AAM14990	Peptide #1424 enco
25	51	43.6	557	22	AAM27425	Peptide #1462 enco
26	51	43.6	557	22	AAM02717	Peptide #1399 enco
27	51	43.6	557	23	ABG36790	Human peptide enco
28	51	43.6	1160	21	AAB23367	Mouse trophinin pr
29	49	41.9	251	18	AAW20965	H. pylori secreted
30	48	41.0	160	22	AAM93460	Human polypeptide,
31	48	41.0	537	22	ABB58083	Drosophila melanog
32	48	41.0	545	22	ABB66603	Drosophila melanog
33	48	41.0	653	23	ABB97943	Human protein sequ
34	48	41.0	829	23	AAU74629	Oestrogen-regulate
35	48	41.0	831	22	AAB87345	Human gene 4 encod
36	48	41.0	831	23	ABG65404	Human albumin fusi
37	48	41.0	831	23	ABB97347	Novel human protei
38	47	40.2	82	21	AAB23377	Human trophinin th
39	47	40.2	86	17	AAR94899	Human trophinin ex
40	47	40.2	86	21	AAB23366	Human trophinin ex
41	47	40.2	671	22	ABG20586	Novel human diagno
42	46.5	39.7	471	14	AAR43646	Environmental adap
43	46	39.3	652	23	ABP27600	Streptococcus poly
44	45.5	38.9	250	21	AAG18354	Arabidopsis thalia
45	45.5	38.9	250	21	AAG45276	Arabidopsis thalia

ALIGNMENTS

RESULT 1  
AAB23358  
ID AAB23358 standard; Protein; 674 AA.  
XX  
AC AAB23358;  
XX  
DT 09-JAN-2001 (first entry)  
XX  
DE Human trophinin protein from residue 69 to 749.  
XX  
KW Trophinin; vaccine; contraceptive; mouse; human; tandem repeat.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Region 1..674  
FT /note= "Tandem repeats"  
XX  
PN US6111089-A.  
XX  
PD 29-AUG-2000.  
XX  
PF 28-FEB-1997; 97US-0808599.  
XX  
PR 04-OCT-1994; 94US-0317522.  
PR 12-MAY-1995; 95US-0439818.  
XX  
PA (BURN-) BURNHAM INST.  
XX  
PI Fukuda MN;  
XX  
DR WPI; 2000-586380/55.

XX Novel nucleic acid sequence encoding mouse trophinin, for producing  
PT trophinin vaccine useful to prevent pregnancy in an individual by  
PT eliciting an immune response which blocks embryo implantation -  
XX  
PS Disclosure; Column 47-52; 57pp; English.  
XX  
CC The present invention relates to mouse trophinin sequence (see  
CC AAB23367). Mouse trophinin is used as a vaccine for preventing  
CC pregnancy in an individual by eliciting an immune response against  
CC trophinin that blocks embryo implantation. Due to the limited  
CC expression of trophinin, an anti-trophinin immune response is less  
CC toxic than other prior art anti-pregnancy vaccines directed to peptide  
CC hormones. The present sequence is the human trophinin protein from  
CC residue 69 to 749. This region of the protein contains tandem  
CC repeats of a 10 residue sequence.  
XX  
SQ Sequence 674 AA;  
Query Match 47.0%; Score 55; DB 21; Length 674;  
Best Local Similarity 50.0%; Pred. No. 9.7;  
Matches 11; Conservative 1; Mismatches 10; Indels 0; Gaps 0;  
QY 2 GVSISAGYQNGFTGNITTSAGF 23  
Db 446 GFSFGNGLSTGFGGLNTSAGF 467  
RESULT 2  
\*ABG13155  
ID ABG13155 standard; Protein; 305 AA.  
XX  
AC ABG13155;  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #13146.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR N-PSDB; AAS77342.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 20; SEQ ID No 43514; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC

CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 305 AA;  
Query Match 46.2%; Score 54; DB 22; Length 305;  
Best Local Similarity 56.2%; Pred. No. 5.7;  
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
QY 7 AGYQNGFTGNITTSAG 22  
Db 240 SGYQGGFTGTVMQTAG 255  
RESULT 3  
ABG08885  
ID ABG08885 standard; Protein; 306 AA.  
XX  
AC ABG08885;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #8876.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR N-PSDB; AAS73072.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 20; SEQ ID No 39244; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as



CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG0010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX

SQ Sequence 306 AA;  
Query Match 46.2%; Score 54; DB 22; Length 306;  
Best Local Similarity 56.2%; Pred. No. 5.7;  
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 7 AGYQNGFTGNITTSAG 22  
:|||||: :||  
Db 241 SGYQGGFTGTVMQTAG 256

RESULT 4  
ABG13158  
ID ABG13158 standard; Protein; 314 AA.

XX AC ABG13158;

XX DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #13149.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
\*KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR N-PSDB; AAS77345.

XX PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -

XX PS Claim 20; SEQ ID No 43517; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating

CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG0010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX

SQ Sequence 314 AA;

Query Match 46.2%; Score 54; DB 22; Length 314;  
Best Local Similarity 56.2%; Pred. No. 5.9;  
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 7 AGYQNGFTGNITTSAG 22  
:|||||: :||  
Db 161 SGYQGGFTGTVMQTAG 176

RESULT 5  
ABG08884

ID ABG08884 standard; Protein; 478 AA.

XX AC ABG08884;

XX DT 13-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #8875.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR N-PSDB; AAS73071.

XX PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -

XX PS Claim 20; SEQ ID No 39243; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 478 AA;  
  
Query Match 46.2%; Score 54; DB 22; Length 478;  
Best Local Similarity 56.2%; Pred. No. 9.4;  
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
  
QY 7 AGYQNGFTGNITTSAG 22  
:|||||: :||  
Db 138 SGYQGGFTGTVMQTAG 153  
  
RESULT 6  
ABG13156  
ID ABG13156 standard; Protein; 478 AA.  
XX  
AC ABG13156;  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #13147.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR N-PSDB; AAS77343.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 20; SEQ ID No 43515; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 478 AA;  
  
Query Match 46.2%; Score 54; DB 22; Length 478;  
Best Local Similarity 56.2%; Pred. No. 9.4;  
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
  
QY 7 AGYQNGFTGNITTSAG 22  
:|||||: :||  
Db 138 SGYQGGFTGTVMQTAG 153  
  
RESULT 7  
ABG08883  
ID ABG08883 standard; Protein; 822 AA.  
XX  
AC ABG08883;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #8874.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR N-PSDB; AAS73070.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 20; SEQ ID No 39242; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human

CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 822 AA;  
  
Query Match 46.2%; Score 54; DB 22; Length 822;  
Best Local Similarity 56.2%; Pred. No. 17;  
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
  
QY 7 AGYQNGFTGNITTSAG 22  
:|||||: :||  
Db 757 SGYQGGFTGTMQTAG 772  
  
RESULT 8  
ABG24362  
ID ABG24362 standard; Protein; 822 AA.  
XX  
AC ABG24362;  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #24353.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
WPI; 2001-639362/73.  
DR N-PSDB; AAS88549.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 20; SEQ ID No 54721; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed

CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 822 AA;  
  
Query Match 46.2%; Score 54; DB 22; Length 822;  
Best Local Similarity 56.2%; Pred. No. 17;  
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
  
QY 7 AGYQNGFTGNITTSAG 22  
:|||||: :||  
Db 757 SGYQGGFTGTMQTAG 772  
  
RESULT 9  
ABG29163  
ID ABG29163 standard; Protein; 822 AA.  
XX  
AC ABG29163;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #29154.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
WPI; 2001-639362/73.  
DR N-PSDB; AAS93350.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 20; SEQ ID No 59522; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 822 AA; Query Match 46.2%; Score 54; DB 22; Length 822; Best Local Similarity 56.2%; Pred. No. 17; Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 7 AGYQNGFTGNITTSAG 22 :||| |||| : :|| Db 757 SGYQGGFTGTVMQTAG 772 :||| |||| : :||

RESULT 10 ABG18082. ID ABG18082 standard; Protein; 861 AA. XX AC ABG18082; XX DT 18-FEB-2002 (first entry) XX DE Novel human diagnostic protein #18073. XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder. XX OS Homo sapiens. XX PN WO200175067-A2. XX PD 11-OCT-2001. XX PF 30-MAR-2001; 2001WO-US08631. XX PR 31-MAR-2000; 2000US-0540217. XX PR 23-AUG-2000; 2000US-0649167. XX PA (HYSE-) HYSEQ INC. XX PI Drmanac RT, Liu C, Tang YT; XX PI WPI; 2001-639362/73. XX DR N-PSDB; AAS82269. XX PT New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity - XX PS Claim 20; SEQ ID No 48441; 103pp; English. XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. CC The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention. CC Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 861 AA; Query Match 46.2%; Score 54; DB 22; Length 861;

Query Match 46.2%; Score 54; DB 22; Length 861; Best Local Similarity 56.2%; Pred. No. 18; Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 7 AGYQNGFTGNITTSAG 22 :||| |||| : :|| Db 796 SGYQGGFTGTVMQTAG 811 :||| |||| : :||

RESULT 11 ABG22537. ID ABG22537 standard; Protein; 861 AA. XX AC ABG22537; XX DT 18-FEB-2002 (first entry) XX DE Novel human diagnostic protein #22528. XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder. XX OS Homo sapiens. XX PN WO200175067-A2. XX PD 11-OCT-2001. XX PF 30-MAR-2001; 2001WO-US08631. XX PR 31-MAR-2000; 2000US-0540217. XX PR 23-AUG-2000; 2000US-0649167. XX PA (HYSE-) HYSEQ INC. XX PI Drmanac RT, Liu C, Tang YT; XX PI WPI; 2001-639362/73. XX DR N-PSDB; AAS86724. XX PT New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity - XX PS Claim 20; SEQ ID No 52896; 103pp; English. XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. CC The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention. CC Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 861 AA; Query Match 46.2%; Score 54; DB 22; Length 861;



DE Human trophinin protein.  
XX Trophinin; vaccine; contraceptive; mouse; human.  
KW Homo sapiens.  
OS US6111089-A.  
XX 29-AUG-2000.  
XX 28-FEB-1997; 97US-0808599.  
XX 04-OCT-1994; 94US-0317522.  
XX 12-MAY-1995; 95US-0439818.  
XX (BURN-) BURNHAM INST.  
XX Fukuda MN;  
XX WPI; 2000-586380/55.  
XX N-PSDB; AAA91964.  
XX Novel nucleic acid sequence encoding mouse trophinin, for producing trophinin vaccine useful to prevent pregnancy in an individual by eliciting an immune response which blocks embryo implantation -  
XX Claim 7; Column 43-48; 57pp; English.  
XX The present invention relates to mouse trophinin sequence (see AAB23367). Mouse trophinin is used as a vaccine for preventing pregnancy in an individual by eliciting an immune response against trophinin that blocks embryo implantation. Due to the limited expression of trophinin, an anti-trophinin immune response is less toxic than other prior art anti-pregnancy vaccines directed to peptide hormones. The present sequence is the human trophinin protein.  
XX Sequence 749 AA;  
SQ Query Match 45.3%; Score 53; DB 21; Length 749;  
Best Local Similarity 39.3%; Pred. No. 22;  
Matches 11; Conservative 7; Mismatches 4; Indels 6; Gaps 1;  
QY 2 GVSISAGY-----QNGFTGNITTSAGF 23  
|::|||: ||:|::||:|  
Db 530 GLNTSAGFGGGLGTSAGFSGGLSTSSGF 557  
RESULT 14  
AAM94034  
ID AAM94034 standard; Protein; 814 AA.  
XX AAM94034;  
AC AAM94034;  
XX 13-NOV-2001 (first entry)  
XX Human stomach cancer expressed polypeptide SEQ ID NO 139.  
DE Human; stomach cancer; marker; screening; micro-metastasis;  
KW peritoneal dissemination.  
XX Homo sapiens.  
OS WO200109317-A1.  
XX 08-FEB-2001.  
XX 28-JUL-2000; 2000WO-JP05063.  
XX 29-JUL-1999; 99JP-0248036.  
XX 27-AUG-1999; 99JP-0300253.  
XX 18-OCT-1999; 99US-0159590.  
XX 11-JAN-2000; 2000JP-0118776.  
XX 17-FEB-2000; 2000US-0183322.

PR 02-MAY-2000; 2000JP-0183767.  
PR 09-JUN-2000; 2000JP-0241899.  
XX (HELI-) HELIX RES INST.  
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T, Aburatani H;  
PI Kodama T, Midorikawa Y;  
XX WPI; 2001-570287/64.  
XX N-PSDB; AAI93912.  
XX New Stomach cancer-associated genes, useful as markers in blood tests for screening for the early stages of the disease -  
XX Claim 1; Page 221-223; 242pp; Japanese.  
XX The invention relates to stomach cancer-expressed genes (AAI93842-AAI93917) and the encoded proteins (AAM93967-AAM94039). The genes can be used as markers in blood tests for screening for the early stages of the disease. The proteins and peptides can be used as targets for screening for compounds to treat the disease. They can also be used for predicting micro-metastases. The gene can predict peritoneal dissemination.  
XX Sequence 814 AA;  
SQ Query Match 45.3%; Score 53; DB 22; Length 814;  
Best Local Similarity 39.3%; Pred. No. 24;  
Matches 11; Conservative 7; Mismatches 4; Indels 6; Gaps 1;  
QY 2 GVSISAGY-----QNGFTGNITTSAGF 23  
|::|||: ||:|::||:|  
Db 595 GLNTSAGFGGGLGTSAGFSGGLSTSSGF 622  
RESULT 15  
AAB94744  
ID AAB94744 standard; Protein; 814 AA.  
XX AAB94744;  
AC AAB94744;  
XX 26-JUN-2001 (first entry)  
XX Human protein sequence SEQ ID NO:15790.  
DE Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
KW Homo sapiens.  
OS EP1074617-A2.  
XX 07-FEB-2001.  
XX 28-JUL-2000; 2000EP-0116126.  
XX 29-JUL-1999; 99JP-0248036.  
XX 27-AUG-1999; 99JP-0300253.  
XX 11-JAN-2000; 2000JP-0118776.  
XX 02-MAY-2000; 2000JP-0183767.  
XX 09-JUN-2000; 2000JP-0241899.  
XX (HELI-) HELIX RES INST.  
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI; 2001-318749/34.  
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs -



XX

PS Claim 8; SEQ ID 15790; 2537pp + CD ROM; English.

XX

CC The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.

XX

SQ Sequence 814 AA;

Query Match

45.3%; Score 53; DB 22; Length 814;

Best Local Similarity 39.3%; Pred. No. 24;

Matches 11; Conservative 7; Mismatches 4; Indels 6; Gaps 1;

OY

2 GVSISAGY-----QNGFTGNITTSAGF 23

|::|||: ||:|::||:|

Db 595 GLNTSAGCGGLGTSAGFSGGLSTSSGF 622

Search completed: July 30, 2003, 16:47:49  
Job time : 85 secs



